DUPLICATE ALSO

Forecasting Research

Forecasting Research Division
Scientific Paper No. 48

New mathematical developments in atmosphere and ocean dynamics, and their application to computer simulations.

by

M. J. P. Cullen

September 1997

Meteorological Office
London Road
Bracknell
Berkshire
RG12 2SZ
United Kingdom

ORGS UKMO F

National Meteorological Library
FitzRoy Road, Exeter, Devon. EX1 3PB

Forecasting Research
Scientific Paper No. 48

New mathematical developments in atmosphere and ocean dynamics, and their application to computer simulations.

by

M.J.P. Cullen

September 1997

Numerical Weather Prediction
Meteorological Office
London Road
Bracknell
Berkshire
RG12 2SZ
United Kingdom

© Crown Copyright 1997

Permission to quote from this paper should be obtained from the above Meteorological Office division.

This chapter has been prepared for the proceedings of the Isaac Newton Institute Programme on Atmosphere and Ocean Dynamics, held in Cambridge from July to December 1996.

New mathematical developments in atmosphere and ocean dynamics, and their application to computer simulations.

M.J.P. Cullen,
Numerical Weather Prediction Development Division,
Meteorological Office,
London Road,
Bracknell, Berkshire RG12 2SZ, U.K.

Abstract

This chapter reviews new developments in the mathematical theory of the partial differential equations which govern the large scale behaviour of the atmosphere and ocean. It then discusses how these can be applied to the models which are used to predict weather and climate, and to define the initial state of the atmosphere or ocean from limited observations. Particular topics include: i) the use of Lagrangian conservation properties in establishing the existence of solutions to and properties of simpler equations which describe the atmosphere and ocean in the limits of rapid rotation and/or strong stratification; ii) demonstrations that the complete equations can be proved to have solutions close to those of the simpler equations in these asymptotic limits; and iii) how the statistical effect of small scale motions interacts with the large scale solutions. It is demonstrated how these techniques can be exploited in computer models suitable for operational applications, and in better ways of exploiting observed data.

1 Introduction

Weather forecasting and climate modelling have advanced tremendously since the pioneering studies of Richardson (1922) and Charney, Fjortoft and von Neumann (1950) first proposed that it was possible to simulate and predict the behaviour of the ocean and the atmosphere by direct computer solution of the classical equations of dynamics and thermodynamics. Since then, computers have been developed which can run operational forecast programmes at 10^{11} floating point operations per second, and much faster in research mode, enabling real time forecasts of the global atmosphere to be made in a few minutes per simulated day on a grid of points about 60km apart in the horizontal and 500m in the vertical. While in the early days of atmosphere and ocean modelling, lack of computer power was clearly the main limiting factor in the quality of the simulations, it is now becoming clear that greater knowledge of the phenomena being simulated should also be exploited in the design of the computer models. Similarly, it is necessary to exploit to the maximum the information contained in observations of the atmosphere and ocean, since the cost of such observations is increasing and the maintenance of even the current level of observations is difficult.

It was therefore timely to hold a six month study programme at the Isaac Newton Institute for Mathematical Sciences in Cambridge to review and discuss the knowledge that mathematicians could bring to bear on meteorological and oceanographic problems, and to advertise to mathematicians the wide scope for quite rigorous studies of atmosphere and

ocean dynamics. This book is a summary of material presented, discussed, and developed during the programme.

2 Flow regimes in the atmosphere and ocean

We start from the basic equations for the dynamics and thermodynamics of an ideal fluid. We write them in the form used by Shutts and Cullen (1987). These are fully compressible, and allow the axis of rotation not to be parallel to the gravitational force. We use Cartesian coordinates to simplify the presentation, but the inclusion of the general direction of the axis of rotation allows immediate application of the results in spherical geometry. The equations are as follows:

$$\begin{aligned} \frac{D\mathbf{v}}{Dt} + 2\Omega \times \mathbf{v} + \nabla\Phi + \alpha\nabla p &= 0 \\ \frac{D\alpha}{Dt} &= \alpha\nabla.\mathbf{v} \\ C_v\frac{DT}{Dt} + p\frac{D\alpha}{Dt} &= 0 \\ p\alpha &= RT \end{aligned} \tag{1}$$

where $\mathbf{v}$ is the wind vector, α the specific volume, p the pressure, C_v the specific heat at constant volume and T the temperature. Ω is the rotation vector, and $\nabla\Phi$ a specified geopotential force, representing gravitational and centrifugal forces. The temperature equations can be replaced by the single equation

$$\frac{D\theta}{Dt} = 0 \tag{2}$$

where θ is the 'potential temperature', which is a function of the entropy and proportional to $p\alpha^\gamma$, where γ is the ratio of specific heats. In the atmospheric context, it is most useful to consider periodic boundary conditions in x, y and a lower boundary in z, with no normal flow across the lower boundary. At the upper boundary we have $p \to 0$ as $z \to \infty$.

The equivalent equations for the ocean are

$$\begin{aligned} \frac{D\mathbf{v}}{Dt} + 2\Omega \times \mathbf{v} + \nabla\Phi + \alpha\nabla p &= 0 \\ \nabla.\mathbf{v} &= 0 \\ \frac{DT}{Dt} &= 0 \\ \frac{DS}{Dt} &= 0 \\ \alpha &= \alpha(S,T) \end{aligned} \tag{3}$$

Here, S represents the salinity. The appropriate boundary conditions are to have a basin Γ with depth $h(x,y)$ and rigid boundaries. There is no normal flow on all boundaries except the upper boundary which is a free surface $z = \eta(x,y) \simeq 0$.

Energy conservation for (1) is expressed as

$$E = \int_\tau \{\frac{1}{2}|\mathbf{v}|^2 + \Phi + C_vT\}\alpha^{-1}d\tau = \text{a constant} \tag{4}$$

These equations describe the complete motion of the atmosphere and ocean. When used for weather and climate forecasting, they have to be averaged in space and time to make computer solution practical. In the atmosphere, the molecular viscous scale is about 10^{-6}m, and in the ocean 10^{-5}m. High resolution global atmospheric models use an averaging scale of at least 50km in the horizontal and about 1km in the vertical, less near the Earth's surface. Models used for limited area forecasting have much shorter averaging scales, down to around 1km. In the ocean, the highest resolution global models currently used have an averaging scale of about 25km in the horizontal. The resolutions used in practice are thus short by a factor of $10^9 - 10^{11}$ of those that would be required for accurate direct solutions. Success in practice depends on being able to define the effect of unresolved motions in terms of the resolved motions, i.e. correct sub-grid modelling (see Pielke (1984) for a review applied to weather forecasting models). This would be much easier if there was a 'spectral gap', i.e. a range of scales where the energy in atmospheric motions was much less than in larger or smaller scales. Observational evidence (Fig. 1) shows a gradual change in spectral slope from k^{-3} at large scales to $k^{-5/3}$ at small scales, but no gaps. There will however be a minimum in the enstrophy (mean squared vorticity) spectrum associated with this change in slope.

The effect of this averaging for the atmosphere is illustrated in Fig.2. This shows the typical space (l) and time (τ) scales associated with different atmospheric phenomena. The line given by $\tau = l^{2/3}$ associated with the energy spectrum $k^{-5/3}$ is plotted along the diagonal. The phenomena occuring along this diagonal are those traditionally associated with 'weather'. Other motions such as internal gravity waves are elsewhere on the diagram. Fig.2 illustrates that there is a difficulty associated with simply averaging the equations in space and time, in that any choice of averaging scale will cut across the peak scales of some phenomena, resulting in them being partly resolved and thus inaccurately represented. If the averaging scale is reduced, then these phenomena may be well predicted, but a new set will become partly resolved. We can also see this from atmospheric observations. Fig. 3 shows two satellite pictures for a similar time. The first covers the North Atlantic and Western Europe. It shows large scale organised phenomena, such as the waving cloud bands extending south-westwards into the Atlantic from Ireland, and regions of highly structured small scale flows, such as south of Iceland. If we zoom in on the United Kingdom, new phenomena become visible, such as the small scale wave train running south-west to north-east across the centre of the British Isles.

There is thus a need to choose averaging scales which make sense in terms of phenomena to be simulated. Having done this, it is necessary to ensure that the equations, with the sub-grid model, have solutions which respect this scale. This will usually mean that they stay smooth on the averaging scale in both space and time. Mathematical results are therefore required to prove that this is so for a given sub-grid formulation. Practical success in this approach requires a clear separation in terms of space and time scale between phenomena that will be simulated and those that will be excluded. The numerical solution should use a resolution significantly finer than the averaging scale, to ensure that the model is integrated accurately.

An alternative approach, which avoids the problem of scale separation to some extent, is to choose a 'sub grid model' which restricts the solutions to specific phenomena, though not requiring smoothness on specific scales. This allows treatment of phenomena such as fronts, which have a small space scale in one direction, but have a large scale in other directions. This leads to 'reduced' systems of equations, which have a simpler set of solu-

tions than the full equations, only describing the phenomena desired. Success in this type of modelling depends on showing that the interaction between resolved and unresolved phenomena is weak. This can be achieved by the geometry of the flow, as well as by scale separation.

One effect of the latter approach is that it is necessary to exclude atmospheric or oceanic states which in reality would be unstable to motions not permitted by the reduced equations. These are likely to include, for instance, states which are statically unstable. In the horizontal, inertial stability is often required. Knox (1997) reviews this issue for a wide set of 'reduced' equations. We can therefore assess the usefulness of the reduced equations by studying how widespread such unstable states are in atmospheric or oceanic observations. An example from such a study is shown in Fig. 4, due to Veitch and Mawson (1993), showing automated aircraft wind data from cases between 30N and 30S where the aircraft track crossed regions of straight flow at close to a right angle. The observations are typically 80km apart. The inertial stability condition is most likely to be violated in low latitudes, because the time-scale of inertial instability is $(2\Omega\sin\phi)^{-1}$, where ϕ is the latitude. This is of the order of 1 day at 5 degrees from the equator. The results shows that in only 5 out of 121 cases was the inertial stability condition violated, suggesting that those reduced systems of equations which require inertial stability may be useful at this horizontal scale even in low latitudes.

2.1 Definition of asymptotic regimes

We define the most important asymptotic regimes in the atmosphere and ocean that we will be considering in terms of the following parameters.

$$\begin{aligned} Ro &= U/fL \quad \text{Rossby number} \\ Fr &= U/NH \quad \text{Froude number} \\ B &= Ro/Fr \quad \text{Burger number} \\ Ma &= U/c \quad \text{Mach number} \\ A &= H/L \quad \text{aspect ratio} \end{aligned} \tag{5}$$

where L, H are horizontal and vertical length scales, U is a horizontal velocity scale, f is the Coriolis parameter, N the Brunt-Vaisala requency, and c the speed of sound. For weather systems in the atmosphere, typical values are $L = 10 - 10^4$ km, $H = 1 - 10$ km, $U = 1 - 100\text{ms}^{-1}$, $f = 0 - 10^{-4}\text{s}^{-1}$, $g = 10ms^{-2}$, $N = 10^{-2}\text{s}^{-1}$, and $c = 300ms^{-1}$. In the ocean $U = 0.1 - 1ms^{-1}$. We will use τ for the time scale. A systematic scale analysis of the governing equations in terms of these parameters is set out, for instance, in Holton (1992); and a more detailed summary of the various 'reduced' sets of equations appropriate in these regimes is given in the chapters by White and Warn later in this volume. We summarise only the main points here.

The asymptotic regimes relevant to weather systems and large scale ocean circulations are characterised by long timescales, and thus balances between the forces occurring in (1) and (3). As discussed by Bartello and Thomas (1996), for instance, there is an issue as to whether the Eulerian or Lagrangian time-scale is greater. In the absence of geographically fixed forcing, such as mountains, their arguments suggest that the Lagrangian time scale is greater, especially in regimes where the k^{-3} spectrum holds. If this is so, we have $\tau = L/U$,

where L is treated as a length scale along trajectories. Other length scales may be shorter.

The horizontal extent of the atmosphere and ocean are much greater than the vertical extent, allowing the existence of flow regimes with very small aspect ratio. To allow these to be identified, we write (1) in terms of horizontal and vertical components of momentum, where the direction $\nabla\Phi$ is defined as vertical. Let W be a vertical velocity scale, then $\tau = H/W$. In the atmospheric case, the vertical variations of the thermodynamic quantities p, θ, α are large. Assume these quantities have background scales p_0, θ_0, α_0, related by the equation of state, and scales of horizontal spatial variations p', T', α'. In the vertical momentum equation, we expect the vertical particle acceleration Dw/Dt to be much less than g. Therefore, the vertical pressure gradient term has to be of a size comparable to g, so $\frac{\partial p}{\partial z}$ is of order $g\alpha_0^{-1}$. The equation of state then allows us to estimate $\frac{\partial \alpha}{\partial z}$ and $\frac{\partial \theta}{\partial z}$. Observed large scale atmospheric flows have large variability in $\frac{\partial \theta}{\partial z}$, including regions where it is zero. It is thus appropriate to estimate $\frac{\partial \alpha}{\partial z}$ to balance $\frac{\partial p}{\partial z}$; with $\frac{\partial \theta}{\partial z}$ assumed smaller in magnitude than either. This gives $\frac{\partial \alpha}{\partial z} \simeq g\alpha_0^{\gamma-1}/(\gamma p_0) = g\alpha_0^{\gamma}/c$.

Equations (1) and (2), with scales of individual terms, become

$$
\begin{array}{rcl}
\dfrac{D\mathbf{v}_h}{Dt} & +2(\Omega \times (\mathbf{v}_h, w))_h + \alpha\nabla_h p = 0 & \\
U\tau^{-1} & (fU, fW) \quad \alpha_0 p' L^{-1} & \\
\dfrac{Dw}{Dt} & +2(\Omega \times (\mathbf{u}, w))_z + \dfrac{\partial \Phi}{\partial z} + \alpha\dfrac{\partial p}{\partial z} = 0 & \\
W\tau^{-1} & (fU, fW) \quad g \quad \alpha_0 p' H^{-1} & \\
\dfrac{\partial \alpha}{\partial t} & +(\mathbf{v}.\nabla\alpha)_h + w\dfrac{\partial \alpha}{\partial z} = \alpha\nabla.\mathbf{v} & (6) \\
\alpha'\tau^{-1} & \alpha_0 U L^{-1} \quad w g \alpha_0^{-\gamma}/c & \\
\dfrac{\partial \theta}{\partial t} & +(\mathbf{v}.\nabla\theta)_h + w\dfrac{\partial \theta}{\partial z} = 0 & \\
\theta'\tau^{-1} & \alpha_0 U L^{-1} \quad w N^2 \theta_0 / g & \\
p\alpha^{\gamma} & = (R\theta)^{\gamma} &
\end{array}
$$

Subscript h indicates horizontal components of vectors, subscript z the vertical component.

We first identify a scaling associated with low Mach number. If we assume that $U = L/\tau$ and that the horizontal acceleration is of similar size to the pressure gradient, the ratio of $\frac{\partial \alpha}{\partial t}$ to $\alpha\nabla.\mathbf{v}$ is U^2/c^2 or Ma^2. The continuity equation thus requires the local divergence to be small. In the atmosphere, the appropriate approximate continuity equation and boundary conditions become

$$
\begin{array}{rl}
\nabla.\alpha\mathbf{v} & = 0 \\
\alpha w & = 0 \text{ at } z = 0, \infty
\end{array}
\qquad (7)
$$

Similar equations apply in the ocean. We next note that fU/g is very small, typically $10^{-3} - 10^{-5}$. The dominant terms in the vertical momentum equation are then

$$\nabla\Phi + \alpha\nabla p = 0 \qquad (8)$$

representing hydrostatic balance. Using (7), we can set $U/L = W/H$, so that $W = UA$.

We now consider regimes where, additionally, either the Rossby number or the Froude number are small. Small Rossby numbers are associated with rapid rotation, and small Froude numbers with strong stratification. In order to calculate the Froude number we need the Brunt-Vaisala frequency N. This is given by

$$N^2 = \frac{g}{\theta}\frac{\partial \theta}{\partial z} \tag{9}$$

We classify these regimes according to the Burger number, (5), as illustrated in Fig. 5. We treat the analysis of the flow in each regime separately. It is also useful to consider the horizontal scale as a function of Burger number with H, f and N^2 fixed, so that

$$L = (NH)/(fBu^{1/2}). \tag{10}$$

If there is rotation, but no stratification, then $B = 0$. In the atmospheric case, this corresponds to uniform potential temperature θ_0. The assumption of small Ro gives geostrophic balance, which using the hydrostatic relation (8) as well can be written:

$$2\Omega \times \mathbf{v} + \nabla\Phi + \alpha\nabla p = 0 \tag{11}$$

The condition $\theta = \theta_0$ means that α is a function of p, so that $\alpha\nabla p$ is a gradient. Using (7) and (11), we derive the generalized Taylor-Proudman theorem that $\mathbf{v}$ does not vary in the direction of Ω. Thus the flow becomes two-dimensional, with no component along the axis of rotation. In the atmosphere, this regime only occurs in limited regions, though it can occur in neutrally stratified layers (for instance observations of Karman vortex wakes behind islands). A similar regime occurs in tropical cyclones, though here it is due to rapid system rotation.

If the rotation is much stronger than the stratification, then $B << 1$ and the horizontal length scale is large compared with the Rossby radius NH/f. This is the appropriate regime for the large scale planetary waves in the extra-tropical atmosphere. However, the geometry of the atmosphere as a thin layer on a spherical surface requires that such motions are horizontal, and not normal to the axis of rotation. In this situation it is appropriate to modify the problem (6) by making the 'shallow atmosphere' approximation (seee chapter by White later in this volume). Write the vertical component of the rotation vector as $f\hat{\mathbf{k}}$ and write $\nabla\Phi = (0, 0, -g)$. The other components of the rotation vector are neglected. The equations then become

$$\begin{aligned}
\frac{D\mathbf{v}_h}{Dt} + (-fv, fu) + \alpha\nabla_h p &= 0 \\
-g + \alpha\frac{\partial p}{\partial z} &= 0 \\
\nabla.\alpha\mathbf{v} &= 0 \\
\frac{\partial \theta}{\partial t} + (\mathbf{u}.\nabla\theta)_h + w\frac{\partial \theta}{\partial z} &= 0
\end{aligned} \tag{12}$$

The geostrophic relation (11) becomes

$$(-fv, fu) + \alpha\nabla_h p = 0 \tag{13}$$

Combining the geostrophic relation, (13), with the continuity equation (7) shows that the horizontal velocity is independent of z if θ is uniform. The vertical scale of such flows, H,

will thus tend to be the maximum permitted by the depth of the atmosphere H_0, (around 10km). This regime is thus only robust for horizontal scales large compared with NH_0/f. A typical value of this in the middle latitude atmosphere is 1000km.

The condition $B = O(1)$, so that the horizontal length scale is comparable with the Rossby radius, and $Ro \simeq Fr \ll 1$ is the natural scaling for 'synoptic' motions in the atmosphere and ocean. For instance, as shown in Gill (1982 chapter 13), it is the natural scale for the baroclinic waves which form the essential dynamics of developing weather systems.

Non-rotating, stratified flow corresponds to $Bu = \infty$. If $Fr \ll 1$, then the first equation of (12) becomes

$$\frac{D\mathbf{v}_h}{Dt} + \alpha \nabla_h p = 0 \tag{14}$$

and the two terms must balance. Standard manipulations can then be used to show that $W \simeq FrUH/L$, so that the flow is approximately horizontally non-divergent. The large scale balance condition is simply $\nabla_h p = 0$, so that there is no flow. The natural solutions will have horizontal density surfaces with gravity waves or gravity currents superposed on them.

If $B \gg 1$ we are considering sub-synoptic horizontal scales, or most low latitude circulations. As in the case $B = \infty$, the flow will be approximately horizontally non-divergent, and the vertical velocity will be small. There will therefore be a tendency for variables to become uncoupled in the vertical and the vertical scale to reduce, destroying the condition $B \gg 1$. There are thus doubts about the sustainability of such a regime.

The best example of such a regime is given by the shallow water equations with a large equivalent depth h. These equations can be obtained by averaging (12) over the depth of the atmosphere. The Froude number is then $U/\sqrt{(gh)}$, and the condition $B >> 1$ is satisfied for mid-latitude values of f and parameters appropriate for the depth averaged atmosphere if $L \ll 3 \times 10^6$. This is smaller than typical deep atmospheric circulations. In low latitudes, this becomes $L \ll 3 \times 10^7$, which is almost always satisfied. Since there is a single layer of fluid, the vertical scale cannot be reduced and the regime will be sustained. The shallow water equations are widely regarded as a useful model for the depth averaged behaviour of the atmosphere and ocean. A similar argument holds for a fluid with a small number of deep homogeneous layers with significantly different density, which can be described by a set of shallow water equations. Such models have been widely used for understanding the development of weather systems, but their accuracy in describing the more continuously stratified flows that occur in the real atmosphere is not known.

There are many other important asymptotic regimes in meteorology and oceanography. We refer to the later chapters by White and Warn and the books by Holton and Gill for a more comprehensive introduction.

3 Lagrangian methods of analysis

There is an extensive literature analysing solutions of (1) relevant to meteorology and of (3) relevant to oceanography, e.g. Gill (1982), based primarily on linearisation of the

governing equations. In this book, we emphasise nonlinear analysis. The main nonlinearities in equations (1) and (3) are associated with the advection operator D/Dt. The effect of the advection operator is to perform a 'rearrangement' of the quantity being advected. The mathematical definitions and properties of rearrangements are discussed in the chapter by Burton and Douglas later in this volume. Here we give the basic definitions and some examples of use. Note that the concepts are simpler if the rearrangements are volume-preserving, rather than mass-preserving. This will be the case in the ocean, and is a good approximation for large scale flows in the atmosphere if the volume is defined in terms of a stretched vertical coordinate which allows for a basic state vertical density variation. The latter procedure is illustrated in section 3.5, following Hoskins (1975).

3.1 Basic ideas of rearrangement theory

We first give a simple example of two scalar functions which are rearrangements of each other. Let $f(x) = x$ for each $x \in [0,1]$, and let g be defined by

$$g(x) = \begin{cases} 1-2x & \text{if } x \in [0,1/2], \\ 2x-1 & \text{if } x \in [1/2,1]. \end{cases} \tag{15}$$

For each $0 < \alpha < 1$,

$$\int_0^1 (f(x)-\alpha)_+ dx = \frac{1}{2}(1-\alpha)^2 = \int_0^1 (g(x)-\alpha)_+ dx \tag{16}$$

where the subscript $+$ denotes the positive part of the function. (16) says that the area of the region which is bounded above by the graph of f and bounded below by the line $y = \alpha$ is the same as the area of the corresponding regions of g. This is illustrated in Figure 6: the total shaded areas are the same. If $\alpha \geq 1$, then $\int_0^1 (f(x)-\alpha)_+ dx = 0 = \int_0^1 (g(x)-\alpha)_+ dx$. Combining the above results, $\int_0^1 (f(x)-\alpha)_+ dx = \int_0^1 (g(x)-\alpha)_+ dx$ for each positive α.

We use this characterization of rearrangements to make a formal definition following Eydeland, Spruck and Turkington (1990). Other possible definitions are discussed by Burton and Douglas in a later chapter.

Let $f, g : \Omega \to \mathbb{R}$, where Ω is a bounded subset of $\mathbb{R}^n$, be two non–negative integrable functions i.e $\int_\Omega f(\mathbf{x})d\mathbf{x} < \infty$, $\int_\Omega g(\mathbf{x})d\mathbf{x} < \infty$. f is a *rearrangement* of g if

$$\int_\Omega (f(\mathbf{x})-\alpha)_+ d\mathbf{x} = \int_\Omega (g(\mathbf{x})-\alpha)_+ d\mathbf{x} \tag{17}$$

for each $\alpha > 0$.

Unless f is constant, there is more than one rearrangement of f. We write $\mathcal{R}(f)$ for the set of all functions which are rearrangements of f.

When using rearrangement theory to study the properties of an evolving flow we consider the trajectory mapping that transforms initial positions $\mathbf{x}(\mathbf{a},0)$ of fluid particles $\mathbf{a}$ into positions $\mathbf{x}(\mathbf{a},t)$ at a later time. We write this mapping as $\nu(0,t)$. Any conserved particle property q is transported by the flow, so that $q(\mathbf{x}(\mathbf{a},t),t) = q(\mathbf{x}(\mathbf{a},0),0)$. We write this more concisely as $q(t) = q(0) \circ \nu(0,t)$. If the fluid is incompressible it follows that, for any t, $q(\mathbf{x},t)$ is a rearrangement of the initial data $q(\mathbf{x},0)$. This is because for each particle in a given subset U of the fluid at time t, we can find the original position of the

particles at time zero and the set of these points has the same size as U. It follows that the trajectory mapping is a measure preserving mapping, the definition of which we give below.

Definition A *measure preserving mapping* from Ω to itself is a mapping $\nu : \Omega \to \Omega$ such that for each (measurable) set $U \subset \Omega$, $\mu(\{x : \nu(x) \in U\}) = \mu(U)$, where μ mesaures the size of the set. We must restrict our definition to sets where we can measure the size of the set - this excludes some pathological choices of U. Halmos (1950) shows that this is equivalent to requiring that for every integrable function f,

$$\int_\Omega f \circ \nu d\mu = \int_\Omega f d\mu. \tag{18}$$

We then introduce the idea of a distance on the set of rearrangements, following ideas of Brenier(1990). Suppose g is a rearrangement of f. We seek to define the distance from g to f in $\mathcal{R}(f)$. Suppose that we can find $\overline{f} \in \mathcal{R}(f)$ such that for every $g \in \mathcal{R}(f)$ there exists a measure-preserving mapping $s : \Omega \to \Omega$ such that $g = \overline{f} \circ s$. Then we can find measure-preserving mappings $s_0, s_1 : \Omega \to \Omega$ such that $f = \overline{f} \circ s_0$ and $g = \overline{f} \circ s_1$. Assuming Ω is compact, we can use the methods of Brenier to calculate the 'geodesic distance' between s_0 and s_1 in the set of measure-preserving mappings from Ω to itself. (In some cases there may be no shortest path between s_0 and s_1, so the concept of a 'generalised path' is needed to solve the problem. In such cases there is no set of continuous trajectories which can carry f to g.) The value we obtain is the distance between f and g.

The difficulties which remain with this approach are firstly to determine when such a $\overline{f} \in \mathcal{R}(f)$ exists, and secondly that the factorisation of f and g into the composition of $\overline{f}$ with a measure-preserving mapping may not be unique.

3.2 Application to verification

Rearrangements can be used to compare two functions of spatial variables. Fig.7 shows two maps of rainfall over the U.K. The first is an observed map from the radar network, the second a computer forecast. Let $f(x,y), g(x,y)$ be two respective functions with values equal to the respective rain rates. Then we can characterise the forecast error simply as the L_2 norm of the difference, $\{\int (f-g)^2\}^{1/2}$. It is more informative, however, to calculate separately the position error of the rain, and the 'statistical properties' of the rain, such as the areas covered by moderate or heavy rain. This can be done by first seeking a function f^* which is a rearrangement of f that is as close to g as possible, and then calculating separately $\| f^* - g \|$ and the distance from f to g as discussed in the preceding subsection. In Fig. 7, the first norm will highlight the absence of heavy rain in the forecast fields, the second will show up the position errors such as the erroneous forecast of rainfall over Wales.

3.3 Interpretation of partial differential equations using rearrangements

The ability to generate rearrangements by measure-preserving mappings suggests that they can be used to give a more general interpretation of partial differential equations governing incompressible flow. Consider equations (3). These contain no spatial derivatives except the gradients of Φ and p, the divergence operator applied to $\mathbf{v}$, and the total derivative operator $\frac{D}{Dt}$. The value of $\nabla\Phi$ will be given explicitly, as in (12). There is no explicit evolution equation for p, the incompressibility condition on $\mathbf{v}$ makes up the

number of equations.

A natural way to define a generalised solution of (3) is through the method of characteristics. Assume the equations are being solved in a region Γ with rigid boundary $\partial\Gamma$. (Note that, as written, (3) has a free surface upper boundary, which would requires the procedure to be generalised). The terms in the equations other than the $\frac{D}{Dt}$ operator give rates of change following fluid particles. The condition $\nabla.\mathbf{v} = 0$ and boundary conditions $\mathbf{v}.\mathbf{n} = 0$ on the boundary $\partial\Gamma$ are interpreted as stating that the time evolution of fluid particle positions is a rearrangement within Γ. Then the trajectory mapping $\nu(0,t)$ defined in section 3.1 is a measure-preserving mapping for any t. Solution of the equations depends on showing that there is a unique trajectory map of the fluid over any time interval $(0,t)$, giving the correct rates of change following fluid particles. We will see two examples of this in the following subsections.

Important points about this procedure are

i) There is no requirement for the position map to be a smooth function of the initial map. It may be possible to prove that it is smooth in a particular case. It is thus most useful if there are no spatial derivatives other than the $\frac{D}{Dt}$ operator. The term ∇p has to be determined implicitly, and we will give two interpretations of it applicable to non-smooth position maps in the subsequent sections.

ii) Similarly, there is no need for particles initially in contact with the boundary to stay there. The rigid boundary condition simply requires all particles to stay inside Γ.

iii) The rearrangement property is imposed, we seek a solution where $\mathbf{x}(a,t)$ is in $\mathcal{R}(\mathbf{x}(a,0))$ for all t, and do not consider 'solutions' outside this space.

iv) It should be possible to generalise this procedure to compressible equations or free surface boundary conditions, where we consider the flow map as performing a 'mass' rearrangement.

3.4 Solution of the incompressible equations

We first illustrate the procedure described above with a simplified model problem which can be derived from (1) or (3). We will use this problem extensively in the rest of this chapter. Assume the axis of rotation and the gravitational force are in the same direction, setting $\nabla\Phi = (0,0,-g)$ with g constant. We write $2\Omega = f$ and take f as constant. We discuss some aspects of the case with f variable, as in (12), in section 3.6. We make a form of Boussinesq approximation (Hoskins (1975)), which only incorporates the effect of buoyancy through variations in the conserved potential temperature θ. In the atmospheric context, this allows a transformation of the vertical coordinate, allowing the equations to be written in incompressible form. The derivation is much more straightforward in the oceanic case from (3). Note that in Hoskins (1975) and papers that follow from it, the pressure p which typically decreases with height is replaced by a geopotential ϕ which typically increases with height and the direction of positive vertical coordinate is reversed. We have not done that in this chapter.

We use Cartesian coordinates (x,y,z), and solve in a finite bounded region Γ, with

boundary $\partial\Gamma$. Write the volume element as τ. The equations then become

$$\begin{aligned}
\frac{D\mathbf{v}_h}{Dt} + (-fv, fu) + \nabla_h p &= 0 \\
\frac{Dw}{Dt} + g\theta/\theta_0 + \frac{\partial p}{\partial z} &= 0 \\
\nabla.\mathbf{v} &= 0 \\
\frac{D\theta}{Dt} &= 0 \\
\mathbf{v}.\mathbf{n} &= 0 \text{ on } \partial\Gamma
\end{aligned} \tag{19}$$

(19) includes explicit equations for the evolution of the velocity components and θ. The pressure is determined implicitly by the incompressibility constraint and the boundary conditions. As a first step to solving (19), ignore the pressure gradient term, then (19) can be solved explicitly. Given initial particle positions $(x(0), y(0), z(0))$ and velocities $(u(0), v(0), w(0))$, calculate a trajectory centre (x_c, y_c) for each particle as $(x(0) + v(0)/f, y(0) - u(0)/f, z(0)$. The particle trajectories consist of circles around the trajectory centres, together with a uniformly accelerating motion in the z direction. The solution at time t is then

$$\begin{aligned}
\theta(t) &= \theta(0) \\
x(t) &= x_c + (x(0) - x_c)\cos(ft) + (y(0) - y_c)\sin(ft) \\
y(t) &= y_c + (y(0) - y_c)\cos(ft) - (x(0) - x_c)\sin(ft) \\
z(t) &= z(0) + w(0)t - \frac{1}{2}g\theta(0)t^2/\theta_0 \\
u(t) &= u(0)\cos(ft) + v(0)\sin(ft) \\
v(t) &= v(0)\cos(ft) - u(0)\sin(ft) \\
w(t) &= w(0) - g\theta(0)t/\theta_0
\end{aligned} \tag{20}$$

This solution does not satisfy the continuity equation. Equivalently, the associated mapping $\nu(0, t)$ of initial particle positions $\mathbf{x}(0)$ to later positions $\mathbf{x}(t)$ is not measure-preserving. To obtain a solution of (19), ν must be projected onto the set of measure-preserving mappings. If we use the 'polar decomposition' theorem of Brenier (1991), we can write

$$\nu = \nabla\chi \circ s \tag{21}$$

where χ is a convex function and s a measure-preserving mapping from Γ to itself. Note that this decomposition only exists under certain restrictions. Burton and Douglas (1997) have reduced these restrictions compared with those required by Brenier. Brenier shows that s is as close to ν in L^2 as any other measure-preserving mapping $\Gamma \to \Gamma$. If ν is non-degenerate, we can then write $s = \nabla\chi^* \circ \nu$. In order to obtain a solution of (19), s must be identified with a continuous trajectory linking initial and final particle positions. As discussed in section 3.1, this is not possible in general. Some of the difficulties are caused by non-uniqueness of the decomposition (21) and degenaricies of ν. It is therefore necessary to seek solutions by splitting the time interval $(0, t)$ into n shorter steps, and replacing g by a sequence $g_n \circ g_{n-1} \circ .... \circ g_1$. If it can be proved that this sequence converges, then the result will be a generalised solution of (20). This has not yet been achieved,

though such a procedure underlies many standard and successful numerical methods.

If g is identified with a continuous trajectory, then particle positions will have to vary smoothly in time. This is a minimum requirement for the solution to make physical sense. However, there is no reason why the resulting velocity field should vary smoothly in space. This gives some plausibility to the idea that typical length scales along trajectories may be greater than length scales in other directions and is consistent with the results of Bartello and Thomas quoted above which were based on Fourier expansions.

Brenier (1991) discusses the relation between this procedure and a solution procedure relying on the Helmholz decomposition of a general velocity field into an incompressible velocity field and the gradient of a scalar. This decomposition is the linearisation of the polar decomposition about the identity map. Though the polar decomposition uses a convex potential, there is no convexity impied of the scalar in the Helmholz decomposition, and thus no implied restriction on the form of the pressure in the solution of (19).

3.5 Extremisation with respect to rearrangements

The scale analysis of section 2 highlighted the importance of the geostrophic and hydrostatic relations (11). We now show how (11) can be derived by a variational argument based on rearrangements, following Shutts and Cullen (1987). The energy associated with solutions of (1) is given by (4). We seek to minimise this with respect to fluid displacements $\xi = (\xi, \eta, \zeta)$ which satisfy the boundary conditions, with changes to the flow variables being given by

$$\begin{aligned} \delta \mathbf{v} &= -2\Omega \times \xi \\ C_v \delta T + p \delta \alpha &= 0 \\ \delta \alpha &= \alpha \nabla . \xi \\ \delta \Phi &= \xi . \nabla \Phi \end{aligned} \tag{22}$$

These changes are consistent with the evolution equations (1) apart from the 'freezing' of the pressure gradient ∇p. Shutts and Cullen show that the geostrophic and hydrostatic relations (11) are the condition for the energy to be stationary with respect to (22). The condition for a minimum is that

$$\delta^2 E = \xi . \Lambda . \xi \geq 0 \tag{23}$$

where

$$\Lambda = 2\Omega \nabla M - \alpha \nabla s \nabla p \tag{24}$$

Shutts and Cullen show that this corresponds to the condition for a fluid parcel to be stable against displacements in a 'frozen' pressure field. (23) implies the condition

$$\det \Lambda \geq 0 \tag{25}$$

The physical justification for this is the natural model of a stable steady state being an energy minimising state. Increased energy corresponds to oscillations about the steady state. If the basic state is time-dependent, but the timescale is much greater than the period of the oscillations about the stable state, then the argument should remain accurate. In this example, the oscillations all have a frequency greater than 2Ω, so that the

argument requires the geostrophic and hydrostatic basic state to evolve on a timescale much slower than this. If the timescales become comparable, the procedure can still be carried out, but will not give useful results.

States which are stationary points of the energy, but not minima, are unstable to motions with a growth rate greater than $(2\Omega)^{-1}$. The evolution will then not be usefully described by oscillations about such a state. However, it is possible that the instability will be released quickly, so that there is a transition on a timescale less than $(2\Omega)^{-1}$ to a state close to a minimum energy state. For this to happen, the motions which achieve this transition must either allow energy loss, or else not satisfy (22). In practice, both are likely.

To illustrate the methods further, we use the simplified model problem (19) introduced in the previous subsection. The conserved energy for these equations is

$$E = \int_\Gamma \{\frac{1}{2}(u^2 + v^2 + w^2) + g\theta z/\theta_0\} d\tau \tag{26}$$

Shutts and Cullen show that if we minimise the energy subject to displacements satisfying

$$\begin{aligned} \delta\mathbf{v} &= -f\mathbf{k} \times \xi \\ \delta\theta &= 0 \\ \nabla.\xi &= 0 \\ \xi.\mathbf{n} &= 0 \text{ on } \partial\Gamma \end{aligned} \tag{27}$$

then the condition for the energy to be stationary is geostrophic and hydrostatic balance in the form

$$(fv, -fu, -g\theta/\theta_0) = \nabla s \tag{28}$$

for some scalar s. (28) still holds if f is not constant, but subsequent derivations do not. The necessary reformulations are discussed in the next subsection. The condition for a minimum is that the matrix $\mathbf{Q}$ whose components are

$$Q_{i,j} = \frac{\partial^2 S}{\partial x_i \partial x_j} \tag{29}$$

$$S = f^{-2}s + \frac{1}{2}(x^2 + y^2) \tag{30}$$

is positive definite. Define

$$\begin{aligned} X &= x + f^{-1}v \\ Y &= y - f^{-1}u \\ Z &= -g\theta/f^2\theta_0 \\ \mathbf{X} &= (X, Y, Z) \end{aligned}$$

(28) shows that

$$(X, Y, Z) = \nabla S \tag{31}$$

(27) thus implies $\delta X = \delta Y = \delta Z = 0$. The solution of the minimisation problem can be interpreted geometrically as finding a convex surface S whose gradients $\mathbf{X}$ take prescribed ranges of values on sets of specified measure. The condition that (29) is positive definite is equivalent to convexity of S. A simple example is the problem of finding a polyhedral surface with faces with specified gradients and areas (Fig.8), corresponding to choosing piecewise constant values of (X, Y, Z) in (31). A proof that this construction is possible for any finite set of values is given in Cullen and Purser (1984). A more general set of such results on constructing polyhedral surfaces is given by Pogorelov (1973).

In general, we can specify the problem as to find $S(x, y, z)$ over a specified region Γ in (x, y, z) given

$$\rho = \det \frac{\partial(x, y, z)}{\partial(X, Y, Z)} \tag{32}$$

$$\int \rho d\mathbf{X} = \mu(\Gamma)$$

where $\mu(\gamma)$ is the volume of Γ. The proof that this problem can be solved is a result of theorem 1.1 of Brenier (1991), susbequently extended by McCann, which shows that a general mapping f from a bounded subset of $\mathbb{R}^n$ to $\mathbb{R}^n$ has a unique rearrangement equal to the gradient of a convex function. In particular it should be noted that S can have discontinuous gradients, even when ρ is a smooth function. Cullen and Purser proposed this as a simple model of atmospheric fronts. Extension of this identification of the problem to periodic domains and shallow water models is given in Cullen and Purser (1989).

3.6 Relation to the Monge-Ampere equation and Monge mass transport problem

Cullen and Purser (1989) showed that (31) has a dual form

$$(x, y, z) = \nabla_{\mathbf{X}} R \tag{33}$$

$$R = \mathbf{x}.\mathbf{X} - S$$

This results from the identification of the transformation from $\mathbf{x}$ to $\mathbf{X}$ as a Legendre transformation (see the chapter by Sewell later in this volume). Using this, (32) becomes

$$\rho = \det \frac{\partial^2 R}{\partial(X, Y, Z)^2} \tag{34}$$

This is a Monge-Ampere equation. Brenier (1991) shows that his 'polar factorization theorem' is equivalent to an existence proof for a generalised solution of (34), subject to the compatibility and boundary conditions

$$\int \rho d\mathbf{X} = \mu(\Gamma) \tag{35}$$

$$\nabla R \in \Gamma$$

The theory of this problem, sometimes called the 'assignment problem' when it arises in other contexts, has been taken much further, see Caffarelli (1996).

If the vertical contribution to the kinetic energy is ignored (it is not varied under (27)), the energy (26) can be written as

$$E = f^2 \int_\Gamma \{\frac{1}{2}((x-X)^2 + (y-Y)^2) - zZ\} d\tau \tag{36}$$

Noting that $\int_\Gamma z^2 + Z^2 d\tau$ is conserved under rearrangements of $\mathbf{X}$, we can study the equivalent problem of minimising

$$E = \frac{1}{2} f^2 \int_\Gamma \{|\mathbf{X}(\mathbf{x}) - \mathbf{x}|^2\} d\tau \tag{37}$$

for all $\mathbf{X} \in \mathcal{R}(\mathbf{X}_0)$, for a given $\mathbf{X}_0$. This may be rewritten as a Monge mass transport problem as follows. Define $\nu(S) = \mu(\mathbf{X}_0^{-1}(S))$ for all subsets of $I\!R^3$ which are sufficiently well behaved. Rewrite (37) as

$$\inf_{s \in S} \int_\Gamma c(\mathbf{x}, s(\mathbf{x})) d\tau \tag{38}$$

where S is the set of measure preserving mappings between (Γ, τ) and $(I\!R^3, \nu)$, and the cost function $c : I\!R^3 \times I\!R^3 \to I\!R$ is defined by $c(\mathbf{x}, \mathbf{y}) = |\mathbf{x} - \mathbf{y}|^2$. Douglas (1996) showed that the infimum in (38) is uniquely attained by a measure-preserving mapping equal to the gradient of a convex function. For a general cost function, conditions under which these problems can be solved are discussed by Gangbo and McCann (1997). Then Cullen et al. (1997) show that a more general form of this problem can be used to extend the theory to the surface of a sphere. Assume that (x, y) and (X, Y) represent coordinates on the spherical surface, and z and Z represent coordinates normal to the surface. Define

$$\mu(\mathbf{x}, \mathbf{X})^2 = \inf_{\mathbf{g} \in G} \{\int_0^1 |f(\mathbf{g}(t))| dt\}^2 - zZ \tag{39}$$

where G is the set of paths connecting $\mathbf{x}$ and $\mathbf{X}$, that is $\mathbf{g} \in G$ satisfies $\mathbf{g}(0) = (x, y), \mathbf{g}(1) = (X, Y)$. Then Cullen et al. show that minimising

$$\int_\Gamma \mu(\mathbf{x}, \mathbf{X})^2 d\tau \tag{40}$$

is equivalent to minimising the energy (36). In effect this procedure replaces the Euclidean metric $\frac{1}{2}((x-X)^2 + (y-Y)^2)$ by a more general metric on the spherical surface. (40) is exactly of the general Monge transport form (38). McCann (1997, private communication) has proved that this problem can be solved.

3.7 Rearrangements and mixing

The energy minimisation problem of section 3.5 is an case where the problem of minimising a functional with respect to rearrangements of functions can be solved uniquely. However, such results are difficult to prove because the set of rearrangements of a given function is not compact, except in the trivial case of a constant function. There can be infinite sequences of such rearrangements of a given function which do not converge to a limit. An example of this is a sequence of arbitrarily fine-grained rearrangements.

Let $f_0 : [0, 1] \to I\!R$ be defined by

$$f_0(x) = \begin{cases} 0 & \text{if } x \in [0, 1/2], \\ 1 & \text{if } x \in [1/2, 1]. \end{cases} \tag{41}$$

Define, for $n \in \mathbb{N}$,

$$f_n(x) = \begin{cases} 0 & \text{if } x = 0, \\ 0 & \text{if } x \in (m/n, (2m+1)/2n], \\ 1 & \text{if } x \in ((2m+1)/2n, (m+1)/n]. \end{cases} \tag{42}$$

where $m = 0, 1, ..., n-1$. The functions f_3 and f_8 are illustrated in Figure 9. For each $n \in \mathbb{N}$, f_n is equal to zero on a set of length $1/2$, and equal to 1 on a set of length $1/2$, therefore f_n is a rearrangement of f_0 for each $n \in \mathbb{N}$. However, given any $g \in L^2(0,1)$, it may be shown that $\int_\Omega f_n g d\mu \to 1/2 \int_\Omega g d\mu$ as $n \to \infty$, that is f_n converges weakly to the constant function with value $1/2$, which is not a rearrangement of f_0.

In applications where we need to take limits and thus ensure that we work with a compact set, we can use the weak closure of the set of rearrangements of a given function. This is the smallest weakly compact set that contains the set of rearrangements, and thus contains all the weak limits of sequences of rearrangements. Douglas (1994) gave the following characterisation

$$\begin{aligned} \overline{\text{conv}R(f_0)} \;=\; \{f \;\text{non}-\text{negative} \mid & \int_\Omega (f-\alpha)_+ d\mu \le \int_\Omega (f_0-\alpha)_+ d\mu \\ & \text{for each } \alpha > 0, \int_\Omega f d\mu = \int_\Omega f_0 d\mu.\} \end{aligned} \tag{43}$$

where the $+$ subscript denotes taking the positive part of the function. If we define f_0 as in (41) it can be shown that any integrable function $\varphi : [0,1] \to \mathbb{R}$ satisfying $0 \le \varphi(x) \le 1$ for each $x \in [0,1]$, and $\int_0^1 \varphi d\mu = 1/2$, belongs to $\overline{\text{conv } R(f_0)}$. This illustrates that $\overline{\text{conv } R(f_0)}$ may be a large class of functions, in particular it includes the constant value $1/2$ which is certainly not a rearrangement of f_0. In general, all rearrangements are included, as are functions derived by smoothing a rearrangement while preserving the value of the integral. Thus the limit of a sequence of fine grained rearrangements will be a smoothed 'average' function. Physically, including these limit functions can be thought of as allowing for a small but finite viscosity.

4 Basic equations of motion and approximation by balanced models

Equations (1) and (3) describe all possible motions of the atmosphere and ocean. In order to understand large scale weather systems, and equivalent ocean circulations, a generic approach is:

i) Identify an asymptotic regime, and thus small parameter, corresponding to the flows of interest.

ii) Identify a system of equations, based on scale analysis, that is appropriate to that regime. We refer to this as a 'reduced' system.

iii) Prove that this system has solutions, which only contain the flows of interest.

iv) Prove that the solutions of the full equations stay close to that of the 'reduced' system in some norm. The estimate will depend on the small parameter used in the original asymptotic analysis.

This argument can be used, in principle, to extend existence results for reduced systems to those for the full fluid equations. Note that it is very difficult to show how a simple solution can stay close to a complicated one, while the converse as in step (iv) may be much more practicable. This procedure has been used, for instance, to show that in appropriate regimes, solutions of the 3 dimensional Euler equations stay close to those of the 2 dimensional Euler equations, which are known to be well behaved. See Babin et al. (1996) and Marsden et al. (1995).

In applications to weather and climate forecasting, it is also important to recognise situations in which the total flow is not well approximated by a reduced system, but the interactions between the motions described by the reduced system and other motions are weak. Such a situation arises in the internal structure of the ocean. This contains large internal gravity waves, Garrett and Munk (1979), which can be shown to couple weakly to the large scale geostrophic circulation, e.g. Gjaja and Holm (1996). We can describe this situation as follows. Let M be the solution space of equations (1) or (3). Let M_0 be the subspace of M which contains solutions of a reduced system of equations. Define a projection P which maps M to M_0. Represent the solution of (1) or (3) as $U(t)$ and that of a reduced system as $u(t)$. Then, as shown in Fig.10, if we initialise the reduced model with data $u(0) = PU(0)$, the effective forecast error of the reduced system at time t is $u(t) - PU(t)$. This may be much smaller than $u(t) - U(t)$, and it is important to estimate it.

We illustrate in the following subsections a particular approach to constructing reduced systems. Later chapters by McIntyre and Roulstone, Holm, and Warn present alternative approaches.

4.1 Approximation by balanced states

We show how a particular reduced system can be defined as a sequence of 'balanced' states. A balanced state is defined for this purpose as a state which is in geostrophic and hydrostatic balance, and is statically and inertially stable (so satisfies (31) with S convex). As shown in section 2, this is the natural lowest order approximation for small Rossby number. A similar analysis is possible for axisymmetric states (Shutts et al. (1988)).

In order to allow a more complete analysis, we start from the simplified problem (19), rather than the original equations (1) or (3). This avoids the complications caused by compressibility. Given a general solution of (19), we wish to show how it can be approximated by balanced states in the above sense. Represent such a state by values $(u, v, w, \theta) = (u_b, v_b, 0, \theta_b)$. Then, following (26), the energy associated with this state is

$$e = \int_\tau \{\frac{1}{2}(u_b^2 + v_b^2 + g\theta z/\theta_0\} d\tau \tag{44}$$

Define a projection Π of a general state onto a balanced state by using the energy minimisation procedure discussed in section 3.5. Thus we minimise e subject to particle displacements satisfying (27). This gives a state

$$\Pi(u, v, w, \theta) = (u_b, v_b, 0, \theta_b) \tag{45}$$

satisfying (31) with S convex.

It is obviously possible to make many choices of projection Π. Using (27) and (31), we see that the choice (45) preserves ρ as defined by (32), or equivalently

$$Q = \frac{\partial(X, Y, Z)}{\partial(x, y, z)} \tag{46}$$

if treated as a function of $\mathbf{X}$. Q is not preserved as a function of $\mathbf{x}$. We refer to Q as the 'balanced' potential vorticity. Note that smooth solutions of (19) preserve the 'Ertel' potential vorticity

$$q = (f + \zeta).\nabla\theta \tag{47}$$

on particles, where $\zeta = \nabla \times \mathbf{u}$. Q and q agree to first order in ζ/f. Other projections which are based on preserving (47) or closer approximations to it are discussed in the chapters by McIntyre and Roulstone and by White. They all provide a generalisation of normal mode projections, which depend on decomposing general fields into a geostrophic and hydrostatic basic state with linear unbalanced waves superposed on it, see Daley (1997). The latter require linearisation of the governing equations about a uniform reference state at rest.

The evolution equations (19) can be written as evolution equations for $\mathbf{X} = (X, Y, Z)$:

$$\begin{aligned} \frac{D\mathbf{X}}{Dt} + f^{-2}\mathbf{k} \times \nabla p &= 0 \\ \frac{Dw}{Dt} + \frac{\partial p}{\partial z} &= -g\theta/\theta_0 \\ \frac{D\alpha}{Dt} &= 0 \\ \mathbf{v}.\mathbf{n} &= 0 \text{ on } \partial\Gamma \end{aligned} \tag{48}$$

where the continuity equation has been written in Lagrangian form. Imagine that we have computed a solution of (48), with particle positions $\mathbf{x}(t)$, and particle values $\mathbf{X}(t)$. At each time t, we project to the minimum energy state using (45). According to (27), this involves displacing the particles to $\mathbf{x}(t) + \xi(t)$ while preserving their values of X, Y and Z. Write $\frac{D^*}{Dt}$ to express a derivative following the 'minimum energy' particle positions $\mathbf{x} + \xi$, and write the 'velocity' that achieves this as $\mathbf{V} = (U, V, W)$. Then the equations expressing the evolution of this balanced state are

$$\begin{aligned} \frac{D^*\mathbf{X}}{Dt} + (f^{-2}\mathbf{k} \times \nabla p)^* &= 0 \\ \frac{D^*\alpha}{Dt} &= 0 \\ (X, Y, Z) &= \nabla S \end{aligned} \tag{49}$$

The * after the term in ∇p indicates that this term is calculated at $\mathbf{x(t)}$ from the solution of (48), and added to the solutions of (49) at positions $\mathbf{x} + \xi$. This means that

(49) cannot be solved without prior knowledge of the solution of (48). The rearrangement means that $(\nabla p)^*$ is in general no longer the gradient of a scalar. The second equation in (49) implies $\nabla.\mathbf{V} = 0$. The boundary conditions on (27) imply that $\mathbf{V}.\mathbf{n} = 0$ on $\partial\Gamma$. The solution of (49) has the same 'balanced potential vorticity' Q as (48) if Q is regarded as a function of $\mathbf{X}$. However, Q is not conserved on particles under (48) and therefore not under (49) either.

The evolution of e can be calculated by first rewriting (49) in terms of the original variables

$$\begin{aligned} \frac{D^*}{Dt}(u_b, v_b) + (-fV, fU) + (\nabla_h p)^* &= 0 \\ \frac{D\theta}{Dt} &= 0 \\ \frac{D\alpha}{Dt} &= 0 \\ (fv_b, -fu_b, -g\theta/\theta_0) &= \nabla s \end{aligned} \tag{50}$$

Multiply the first of these equations by (u_b, v_b) and the second by z and add. Using $W = \frac{D^*z}{Dt}$, and the fact that $\mathbf{V}.\nabla s = -fUv_b + fVu_b - gW\theta/\theta_0$ integrates to zero, we find that

$$\frac{De}{Dt} = \int (u_b, v_b, 0).(\nabla p)^* \tag{51}$$

This is zero if $(\nabla p)^*$ can be written as $\nabla\pi$ for some scalar π, because $\nabla.(u_b, v_b, 0) = 0$. For flows close to balance, so that $|\xi|$ is small, we can write

$$(\nabla p)^* = \nabla p - \xi.\nabla(\nabla p) + O(|\xi|^2) \tag{52}$$

The condition that $\nabla p^* = \nabla\pi$ becomes

$$\frac{\partial\xi}{\partial y}\frac{\partial^2 p}{\partial x^2} + \frac{\partial\eta}{\partial y}\frac{\partial^2 p}{\partial x \partial y} - \frac{\partial\xi}{\partial x}\frac{\partial^2 p}{\partial y \partial x} - \frac{\partial\eta}{\partial x}\frac{\partial^2 p}{\partial y^2} = 0 \tag{53}$$

This is satisfied for flows independent of one horizontal coordinate. It is of interest to seek other cases where (53) is either zero, or smaller than would be expected from general estimates. In such cases $\frac{\partial e}{\partial t}$ will be small, and if $E - e$ is initially small, it will only grow slowly, so that the flow stays close to balance. In the general case, (51) and (52) allow $E - e$ to be estimated in terms of ξ and p. We will exploit this in the next subsection.

4.2 Approximation of evolution by a reduced set of equations

We now illustrate how a reduced set of equations can be constructed, which only uses knowledge of balanced states, and how estimates of the difference between the solution of a reduced system and the full system can be made.

A reduced set of equations for predicting the evolution of a balanced state, approximating the real state of the atmosphere, can be obtained by replacing the term $(f^{-2}\mathbf{k} \times \nabla p)^*$ in (49) by a term calculated directly from the balanced state variables. The geostrophic and hydrostatic pressure in the balanced state is given by (28) as s. In order to calcuate $(\nabla p)^*$ from the balanced variables, we first need to estimate the true pressure p from s,

and then estimate $(\nabla p)^*$ from (52). However, the latter estimate depends on ξ, which is arbitrary and unrelated to the balanced state. It can only be estimated by first estimating the energy difference $E - e$ between the true state and the balanced state, and then using (23) to estimate ξ. We illustrate two estimates, appropriate for different asymptotic regimes.

Consider first the case $Ro \leq O(1), B = NH/fL \gg 1$ which implies $Fr \ll 1$. As discussed in section 2, we then have $W/H << U/L$ and hydrostatic balance which for equations (19) means $\frac{\partial p}{\partial z} = -g\theta/\theta_0$. The definition of N^2, (9), means that $g/\theta_0 \frac{\partial \theta}{\partial z} \gg (fL/H)^2$, so that there is a large vertical variation of θ. First calculate what vertical θ conserving displacement is needed to change the pressure, assumed hydrostatic, from p to s. This requires a change of θ of magnitude $(p-s)\theta_0/(gH)$ and thus a vertical displacement of magnitude $(p-s)\theta_0/(gH\frac{\partial \theta}{\partial z}) = (p-s)/(N^2H)$. The condition $\nabla.\xi = 0$ means that the associated horizontal displacements will have to be of magnitude $L(p-s)/(N^2H^2)$. Condition (27) then shows that the horizontal velocity components will be changed by an amount $\delta u \simeq fL(p-s)/(N^2H^2)$. If, instead, the horizontal velocity components were changed by a horizontal displacement so as to be in geostrophic balance with the exact pressure p, the change required would be $(p-s)/(fL) \simeq B^2 \delta u$, which is larger by the factor B^2. Therefore, in this regime, the projection Π will change the pressure field to match the horizontal velocity components, rather than vice versa. This is consistent with geostrophic adjustment theory, see Haltiner and Williams (1980). In particular it will preserve the horizontal non-divergent velocity components to $O(B^{-2})$.

We can therefore estimate p to $O(B^{-2})$ by assuming $\nabla \times (u, v, 0) = \nabla \times (u_b, v_b, 0)$. If $Ro = O(1)$, then $Fr \simeq B$ and as shown in section 2 the horizontal divergence is $O(B)$ times the vertical component of vorticity. Then to $O(B)$ we can impose the condition $\nabla.(u, v, 0) = 0$ and hence $w = 0$ in (27). Substituting $\nabla \times (u, v, 0) = \nabla \times (u_b, v_b, 0)$ gives

$$\nabla.(\mathbf{u_b}.\nabla \mathbf{u_b}) + 2\nabla.(-fv_b, fu_b, 0) + \nabla^2 \pi = 0 \tag{54}$$

where $\mathbf{u}_b = (u_b, v_b, 0)$ and π is the resulting estimate of p. An appropriate set of reduced equations is then

$$\begin{aligned} \frac{D}{Dt}(u_b, v_b) + (-fV, fU) + \nabla_h \pi &= 0 \\ \frac{D\theta}{Dt} &= 0 \\ \frac{D\alpha}{Dt} &= 0 \\ (fv_b, -fu_b, g\theta/\theta_0) &= \nabla s \end{aligned} \tag{55}$$

together with (54). Since the pressure gradient term is approximated by the gradient of a scalar, these equations conserve the balanced energy e as defined by (44). In the special case where the initial data for (19) satisfies $\nabla.(u, v, 0) = 0$ and has $\theta = \theta(z)$ only, (55) reproduces the solution of (19) exactly.

The argument above also shows that in the case $B \ll 1$, the projection preserves the pressure field to $O(B^2)$. In this case, we simply set $\pi = s$, so that the appropriate set of reduced equations is

$$\frac{D}{Dt}(u_b, v_b) + (-fV, fU) + \nabla_h s = 0$$

$$\begin{aligned} \frac{D\theta}{Dt} &= 0 \\ \frac{D\alpha}{Dt} &= 0 \\ (fv_b, -fu_b, g\theta/\theta_0) &= \nabla s \end{aligned} \tag{56}$$

These are precisely the semi-geostrophic equations studied by Cullen and Purser (1989) and many others.

This analysis suggests that the semi-geostrophic model is a good model of balanced flow if $B \ll 1$, but (55) is superior if $B \gg 1$, and is exact in the special case of non-divergent flow with stratification independent of the horizontal coordinates. This is consistent with results in the literature. The case $B \gg 1$ corresponds to most published integrations with the shallow water equations. Allen et al. (1990) and others have shown that the semi-geostrophic model is not very accurate in such cases, and a number of other reduced models are better. The chapter by Roulstone and McIntyre later in this volume shows how the replacement of the pressure gradient by an approximation is a generic method of constructing reduced systems of equations.

As shown in the previous subsection, a complete estimate of the difference between the solution of the full equations and reduced equations also requires an estimate of the magnitude of the displacements ξ. This can be deduced from the energy difference $E - e$. Using (23), we obtain

$$E - e \geq \int \lambda_{\min} |\xi|^2 \tag{57}$$

where $\lambda_{\min}$ is the smallest eigenvalue of λ as defined by (24). We then have to estimate $E - e$ in terms of ξ and p. Estimates of p can be obtained in principle from estimates of s and $\mathbf{u_b}$ as in the derivation of (55) and (56). Estimates of s have to be obtained from an existence theory for the chosen reduced system of equations. This programme of work has not yet been carried through.

4.3 Existence of solutions to reduced equations

In this subsection we discuss the properties of the solutions of the two characteristic reduced systems (56) and (55). Equations (56) can be written in terms of $\mathbf{X}$ variables as

$$\begin{aligned} \frac{D\mathbf{X}}{Dt} + f^{-2}\mathbf{k} \times \nabla s &= 0 \\ \frac{D\alpha}{Dt} &= 0 \\ (X, Y, Z) &= \nabla S \end{aligned} \tag{58}$$

Using (29) and (31), the last term in the first equation of (58) has components

$$(-\frac{\partial s}{\partial y}, \frac{\partial s}{\partial x}, 0) = (y - Y, X - x, 0) \tag{59}$$

It was shown by Cullen and Purser (1989) that ρ as defined by (32) is conserved under (58) following fluid particles. We can thus rewrite equations (58) and 59), using (33) and (34), as

$$\frac{D\rho}{Dt} = 0$$
$$\frac{D}{Dt} \equiv \frac{\partial}{\partial t} + \mathbf{U}.\nabla_{\mathbf{X}}$$
$$\mathbf{U} = (y - Y, X - x, 0) \tag{60}$$
$$(x, y, z) = \nabla_{\mathbf{X}} R$$
$$\rho = \det \frac{\partial^2 R}{\partial (X, Y, Z)^2}$$

These equations are to be solved in $\mathbf{X}$ coordinates for all $\mathbf{X} \in I\!R^3$, with boundary condition

$$\mathbf{x} = \nabla R \in \Gamma \tag{61}$$

(60) and (61) can be interpreted as decsribing a motion of particles in (X, Y, Z) space, with velocity $\mathbf{U} = (y - Y, X - x, 0)$, conserving ρ. The solution procedure is therefore as follows. Given $\rho(X, Y, Z)$ as defined by (32), satisfying the compatibility conditions $\rho \geq 0, \int \rho = \mu(\Gamma)$, use the solution procedure discussed in section 3.5 to find $R(X, Y, Z)$. This gives $\mathbf{x}$ as a function of $\mathbf{X}$, and allows $\mathbf{U}$ to be computed. The solution can then be advanced in time. Further details are given in Cullen and Purser (1989).

Note that the 'velocity' $\mathbf{U}$ can be calculated without any differentiation, so that it makes sense almost everywhere for any convex solution R. It is conjectured that the evolution equations (58) then make sense for any convex R, whether differentiable or not. Proving good behaviour of such 'transport' equations requires one of the two following properties to hold, (Brenier, private communication). Either

$$|U(t, \mathbf{X}) - U(t, \tilde{\mathbf{X}})| \leq C(t)\eta(|\mathbf{X} - \tilde{\mathbf{X}}|) \tag{62}$$

where $C \in L^1_{\text{loc}}(I\!R^+)$ and $\int_0^1 \frac{ds}{\eta(s)} = \infty$. This is Cauchy-Lipschitz theory with Osgood's condition (see Gerard (1992)). Or

$$\int_{|\mathbf{X}| \leq R} |\nabla \mathbf{U}(t, \mathbf{X})| dx \leq C_R(t), \forall R < \infty \text{ with } C_R \in L^1_{\text{loc}}(I\!R^+) \tag{63}$$

This corresponds to the transport theory of DiPerna and Lions (1989). Recent results by Caffarelli suggest that the second of these conditions at least can be met.

Now consider equations (55). These can be written in terms of $\mathbf{X}$, giving equations similar to (58) apart from the replacement of the last terms in the first equation by $f^{-1}\mathbf{k} \times \nabla\pi$. This equation no longer conserves ρ following particles, since in general $\nabla\pi$ cannot be written as a gradient in $\mathbf{X}$ space. It will, however, satisfy conservation of a form of potential vorticity closer to (47). In the case $w = 0$, the first equation of (55) is exactly the equation for the evolution of two-dimensional incompressible flow. The theory for these equations is well established, giving existence and regularity under suitable conditions for infinite time, see e.g. Kato and Ponce (1986). In the special case where θ is a function of z only and $\mathbf{u}_b$ a function of (x, y) only, the solution of (55) satisfies $w = 0$ and so exists for all time, and is also a solution of (19). In general, it is necessary to control w, as in the work of Babin et al. (1996) and Marsden et al. (1995). In the other special case where (55) is vertically averaged, to give a reduced form of the shallow water equations, there is a good chance that $\nabla_h.\mathbf{v}$ can be controlled and a theory obtained based on the results for the case $\nabla_h.\mathbf{v} = 0$. See Babin et al. (1997) or Embid and Majda (1996).

4.4 Calculation of solutions to reduced equations

We illustrate first a solution procedure for equations (56). Write them in the form

$$\begin{aligned} \mathbf{Q}\begin{pmatrix} u \\ v \\ w \end{pmatrix} + \frac{\partial}{\partial t}\nabla s &= \mathbf{H} \\ \nabla.\mathbf{u} &= 0 \\ (fv_b, -fu_b, g\theta/\theta_0) &= \nabla s \end{aligned} \tag{64}$$

where

$$\mathbf{Q} = \begin{pmatrix} fv_x + f^2 & fv_y & fv_z \\ -fu_x & f^2 - fu_y & -fu_z \\ g\theta_x/\theta_0 & g\theta_y/\theta_0 & g\theta_z/\theta_0 \end{pmatrix} \tag{65}$$

and

$$\mathbf{H} = \begin{pmatrix} f^2 u_b \\ -f^2 v_b \\ 0 \end{pmatrix} \tag{66}$$

This formulation is essentially due to Schubert (1985). Equations (55) take a similar form, with (66) replaced by

$$\mathbf{H} = \begin{pmatrix} -f\frac{\partial \pi}{\partial y} \\ -f\frac{\partial \pi}{\partial x} \\ 0 \end{pmatrix} \tag{67}$$

Either of these equations can be solved by forming a single equation for $\frac{\partial \nabla s}{\partial t}$:

$$\begin{aligned} \nabla.\mathbf{Q}^{-1}\frac{\partial}{\partial t}\nabla s &= \nabla.\mathbf{Q}^{-1}\mathbf{H} \\ (\mathbf{Q}^{-1}\frac{\partial}{\partial t}\nabla s)_n &= (\mathbf{Q}^{-1}\mathbf{H})_n \text{ on } \partial\Gamma \end{aligned} \tag{68}$$

These form an elliptic equation for $\frac{\partial \nabla s}{\partial t}$ if $\mathbf{Q}$ has no negative eigenvalues. For the semi-geostrophic equations, Shutts and Cullen (1987) show that $\mathbf{Q}$ is exactly the matrix that appears in (29), and negative eigenvalues correspond to unstable states not describable by semi-geostrophic theory.

4.5 Generic forms of reduced equations

Equations (60) are an example of a generic form of reduced equations for large scale atmospheric and oceanic flow which take the form

$$\begin{aligned} \frac{Dq}{Dt} &= 0 \\ \frac{D}{Dt} &\equiv \frac{\partial}{\partial t} + \mathbf{U}.\nabla \\ \mathcal{H}(\mathbf{U}, \theta) &= q \end{aligned} \tag{69}$$

q is a form of potential vorticity such as (46) or (47). $\mathcal{H}$ contains an elliptic operator which allows all the other fields to be derived from q. In the case of (60), this is the Monge-Ampere equation for R, and the calculation of $\mathbf{U}$ from R. Boundary conditions have to be chosen to allow solution of the elliptic problem. (61) are the appropriate conditions for (60), and do not imply any physical restriction other then no flow through rigid boundaries. Other examples where the Ertel form of potential vorticity (47) is used are given in the chapter by Roulstone and McIntyre. Boundary conditions are more of an issue in these cases.

The conservation law $\frac{D\theta}{Dt}$ is implied by (69). The statement is often then made that potential vorticity is advected quasi-two dimensionally along θ surfaces. However, this ignores the nontrivial motion of the θ surfaces themselves, and may be misleading. In particular, the θ surfaces will usually intersect the lower and, in the ocean, upper boundary at positions which change in time.

This formulation is very convenient for proving existence of solutions, as shown in section 4.3. However, it also allows information about the properties of the solution to be deduced. The conservation of potential vorticity makes the equations similar to the equations for two-dimensional incompressible flow, which conserve vorticity. The equations for three-dimensional incompressible flow do not conserve any vorticity-like quantity. Hence the large scale flow of the atmosphere and ocean is often described in terms of two-dimensional turbulence, e.g. Leith (1983).

Equations (69) also allow deductions about the sensitivity of the solutions to small perturbations. If the fluid is displaced by a field χ, retaining potential vorticity conservation, then we have

$$\mathcal{H}(\mathbf{U}+\delta\mathbf{U},\theta+\delta\theta) = q - \chi.\nabla q \tag{70}$$

(70) can be linearised and solved for $(\delta\mathbf{U},\delta\theta)$ in terms of χ. The key quantities will be the eigenfunctions and eigenvalues of $\mathcal{H}^{-1}$. Small eigenvalues will result in large sensitivity of the solutions to perturbations in the direction of the associated eigenfunction, unless $\chi.\nabla q$ happens to be small. These properties are exploited in section 8.5.

The most useful reduced systems of the form (69) also conserve energy. Both (55) and (56) conserve the balanced energy e. As discussed in the later chapters by Roulstone and McIntyre, Holm, and Morrison, such systems can often be written in Hamiltonian form, with the potential vorticity equation expressing the Liouville theorem or conservation of symplectic structure. Where the conserved potential vorticity takes the form (46), the conservation law is clearly seen as a conservation of phase space volume, with $\mathbf{X}$ being the phase space coordinates.

5 Sub-grid models and their interaction with resolved dynamics

5.1 Introduction

Equations (1) and (3) have to be solved in practical applications with a sub-grid model which expresses the effects of unresolved motions. It is also necessary to include forcing

terms. In the atmosphere the effects of phase changes of water are particularly important. This section is primarily written from an atmospheric viewpoint, though the purely dynamical aspects apply to the ocean as well.

The sub-grid model in operational atmospheric models includes
a) Actual forcing terms, such as radiative heating.
b) Explicit terms describing phase changes of water.
c) Interactions with the lower boundary.
d) Averaged effects of sub-grid dynamics.

A generalisation of (19) to include these effects can be written

$$\begin{aligned} \frac{D\mathbf{u}_h}{Dt} + (-fv, fu) + \nabla_h p &= (F_u, F_v) \\ \frac{Dw}{Dt} - g\theta/\theta_0 + \frac{\partial p}{\partial z} &= 0 \\ \nabla.\mathbf{v} &= 0 \\ \frac{D\theta}{Dt} &= F_h + H - LP + S_h \\ \frac{Dr}{Dt} &= F_r + P + S_r \\ \mathbf{v}.\mathbf{n} &= 0 \text{ on } \partial\Gamma \end{aligned} \tag{71}$$

Here r represents the water vapour content, (F_u, F_v, F_h, F_r) represent subgrid increments to (u, v, θ, r) respectively. S_h, S_r are source terms for heat and moisture at the lower boundary. H is a source term for heat in the free atmosphere (typically radiation). P is a source/sink term for water vapour, resulting in a term LP in the thermodynamic equation where L is the latent heat. We only consider very simple versions of these terms, ignoring such issues as the difference between condensed water and ice.

If equations (19) were being solved exactly, only the source terms would have to be included and (F_u, F_v, F_h, F_r) could be omitted. In the realistic situation where the equations have to be averaged in space and time before being solved, it is of interest to study how the sub-grid model interacts with the larger scale dynamics. One way of doing this is to include the same effects in a reduced set of equations, and study the effect on the solutions. Observations and computations both show that reduced systems of equations can be quite accurate in describing weather systems even where there are strong effects due to latent heat release and there are significant regions of instability, such as in the mesoscale convective systems which are common over continents in summer. A recent such study is described by Olsson and Cotton (1997). Similarly, observations show that, if a simple boundary layer model is incorporated in the reduced dynamics, the interactions of fronts with the atmospheric boundary layer can be described well, e.g. Ostdiek and Blumen (1997).

5.2 Effect on energy minimisation

Geostrophic and hydrostatic balance can be generalised for equation (71) to

$$(-fv, fu) + \nabla_h p = (F_u, F_v) \tag{72}$$

$$g\theta/\theta_0 + \frac{\partial p}{\partial z} = \quad 0$$

The first equation in (72) is sometimes called 'geotriptic' balance (Johnson (1966)), and we will use this term forthwith. If (F_u, F_v) can be linearised to a form $-c_D(u, v)$, then we can show that (72) represents the condition for the energy (26) to be made stationary subject to variations

$$\begin{aligned} \delta\mathbf{v} = &\quad f(\eta, -\xi) - c_D(\xi, \eta) \\ \delta\theta = &\quad 0 \\ \nabla.\xi = &\quad 0 \\ \xi.\mathbf{n} = &\quad 0 \text{ on } \partial\Gamma \end{aligned} \tag{73}$$

With this linearisation of F, the geotriptic and hydrostatic relations (72) can be written

$$(fv - c_D u, -fu - c_D v, -g\theta/\theta_0) = \nabla s \tag{74}$$

The condition for a minimum is that the matrix $\mathbf{Q}$ whose components are

$$\mathbf{Q} = \begin{pmatrix} fv_x - c_D u_x + f^2 + c_D^2 & fv_y - c_D u_y & fv_z - c_D u_z \\ -fu_x - c_D v_x & f^2 + c_D^2 - fu_y - c_D v_y & -fu_z - c_D v_z \\ g\theta_x/\theta_0 & g\theta_y/\theta_0 & g\theta_z/\theta_0 \end{pmatrix} \tag{75}$$

is positive definite. Equation (75) shows that the effect of the drag term is to increase the effective stability of the flow and make it easier to satisfy the condition for energy minimisation. Knox (1997) reviews several studies which show that sub-grid scale effects lead to modifications of the inertial stability criteria, and hence allow the existence of regions in the atmosphere that appear to violate the large scale inertial stability condition.

The further manipulations of section 3.5 can only be carried out if we assume c_D is independent of x and y. This is not very realistic, though it would be a viable approximation for process studies, where we study the effect of surface friction on weather systems over a homogeneous surface (e.g. all ocean, or all pine forest).

5.3 Effect on maintenance of balance in the flow

We seek a reduced system of equations based on equations (71). To do this, we make some simple but qualitatively realistic choices for the various sub-grid terms. Greater sophistication is obviously possible.

As in the previous subsection, let $(F_u, F_v) = -c_D(u, v)$. Set $F_h = 0$, and treat S_h as a specified function of position. The moisture source term that affects the thermodynamic equation depends on the rate of change of the difference bwteen the saturation vapour pressure and the actual pressure following a fluid particle. This depends strongly on pressure, which has a large vertical variation. A useful simple approximation is to set $LP = w\frac{\partial}{\partial z}(\theta - \theta_E)$, where θ_E is the 'effective' potential temperature. This should only be used in saturated regions, where $r \geq r_{\mathrm{SAT}}(p)$. Much more information is given in standard meterological textbooks, such as Haltiner and Williams (1980). These simplifications decouple the moisture variable r from the equations, except at the boundary between saturated and unsaturated regions. This choice of LP allows the most important

effect of moisture on the large scale dynamics to be discussed.

We now write down a reduced system of equations analogous to the semi-geostrophic equations (56). In principle, other sets of reduced equations, such as (55), could be used as a starting point. (56) can be written

$$\begin{aligned} \frac{D}{Dt}(u_b, v_b) + f(v_b - V, U - u_b) &= 0 \\ \frac{D\theta}{Dt} &= 0 \\ \frac{D\alpha}{Dt} &= 0 \\ (fv_b, -fu_b, g\theta/\theta_0) &= \nabla s \end{aligned} \tag{76}$$

An extension of these to include the simplified sub-grid terms is

$$\begin{aligned} \frac{D}{Dt}(u_c, v_c) + f(v_c - V, U - u_c) &= -c_D(u_c - U, v_c - V) \\ \frac{D\theta_E}{Dt} &= S_h; r \geq r_{SAT} \\ \frac{D\theta}{Dt} &= S_h; r < r_{SAT} \\ \frac{Dr}{Dt} &= 0 \\ \frac{D\alpha}{Dt} &= 0 \\ (fv_c - c_D u_c, -fu_c - c_D v_c, g\theta/\theta_0) &= \nabla s \end{aligned} \tag{77}$$

The particular form of the first equation is designed to ensure the drag acts as a sink of kinetic energy. It can be shown that a term $-c_D(u_c^2 + v_c^2)$ is added to the kinetic energy equation. (77) can be written in the form (64), where now

$$\mathbf{Q} = \begin{pmatrix} fv_{cx} - c_D u_{cx} + f^2 + c_D^2 & fv_{cy} - c_D u_{cy} & fv_z - c_D u_{cz} \\ -fu_{cx} - c_D v_{cx} & f^2 + c_D^2 - fu_{cy} - c_D v_{cy} & -fu_{cz} - c_D v_{cz} \\ g\Theta_x/\theta_0 & g\Theta_y/\theta_0 & g\Theta_z/\theta_0 \end{pmatrix} \tag{78}$$

$$\mathbf{H} = \begin{pmatrix} (f^2 + c_D^2)u_c \\ -(f^2 + c_D^2)v_c \\ gS_h/\theta_0 \end{pmatrix} \tag{79}$$

with $\Theta = \theta$ if $r < r_{\mathrm{SAT}}$, $\Theta = \theta_E$ otherwise. This simple form of equation is not valid at saturation boundaries.

We can see by comparing (78) and (79) to (64) that the effect of the sub-grid terms is divided between altering the **Q** matrix, which determines the response to the forcing, and altering the forcing itself. Thus friction makes the model less responsive to forcing, and latent heating makes it more responsive. The friction also contributes to the forcing directly, while latent heating does not. It is therefore rather misleading to think of latent heating driving large scale atmospheric circulations, it is better to think of it as changing the characteristic motions that result. See, for instance, Thorpe and Emanuel (1985).

A particular effect of latent heating is through the effect of saturation boundaries, which is not directly treatable in (78). If the air is either saturated or unsaturated everywhere in a region, observations show that negative eigenvalues of $\mathbf{Q}$ can only be sustained where the forcing is very strong, such as in the lower layers of the atmosphere over subtropical deserts. The approximation $\det\mathbf{Q} = 0$ is quite good in frontal zones with active precipitation, Emanuel (1983). The initiation of major convective events, such as thunderstorms, is associated with the rapid transition of a significant mass of air from unsaturated to saturated, which results in a discontinuous loss of stability.

A simplified model of this process has been constructed by Shutts (1987). He solved a finite dimensional version of (77) in a vertical cross section without frictional drag. The data was represented by piecewise constant values of $v_c + fx$, θ and r. All variables were considered as independent of y. The solution procedure described in section 3.5 was used to construct solutions. The effects of latent heat release were included by carrying out standard parcel thermodynamic calculations. An iterative procedure was used, first constructing the solution ignoring latent heating, then calculating the latent heating and solving for new parcel positions. When there is such a discontinuous loss of stability, the parcels jump in the vertical to a new equilibrium position and energy is lost. This can be interpreted as there being a mass sink at low levels and a mass source at high levels, Shutts (1995). The solutions are well-defined without including the effects of evaporative cooling, which according to Emanuel et al (1994) are needed to make the large scale flow stable. The convective transport stabilises the atmosphere by redistributing the heating away from the source region through the jump in the parcel position and the compensating subsidence of other parcels. The lost energy is assumed to be converted into unbalanced waves which disperse and dissipate.

This vertical jump corresponds to a mass rearrangement which cannot be generated by a smooth trajectory, and thus represents a 'generalised flow', as discussed in section 3.1. It should be possible to extend the mathematical theory to cover this case and prove that the above iteration converges. This would involve characterising the solution of (64) with a matrix $\mathbf{Q}$ which is not positive definite as having a solution for (u, v, w) which is a generalised flow, rather than a continuous velocity field. However, the 'pressure tendency' $\frac{\partial s}{\partial t}$ will remain well-defined and bounded everywhere.

Standard procedures for representing moist convection in operational weather forecasting models use this type of iteration, applied in the vertical only. The theory discussed above may allow these procedures to be put on a sound basis. A similar model has been used by Shutts (1987) to discuss the interaction of balanced flow with orography.

5.4 Maintenance of large scale balances by sub-grid transport

In this subsection we consider a more general approach to represnting the effects of the sub-grid dynamical terms (F_u, F_v, F_h, F_r) in (71). This is based on methods developed for the ocean by Gent and McWilliams (1996). A similar analysis for the atmosphere has been carried out by Buhler (1996).

The sub-grid dynamical terms come from averaging the nonlinear terms in the basic equations. If we take these to be (19), together with a moisture conservation equation, then

$$\frac{\overline{D}\overline{\mathbf{v}}_h}{Dt} + (-f\overline{v}, f\overline{u}) + \nabla_h \overline{p} = \overline{\mathbf{v}}.\nabla\overline{\mathbf{v}}_h - \overline{\mathbf{v}.\nabla\mathbf{v}_h}$$

$$\frac{\overline{D}\overline{w}}{Dt} + g\overline{\theta}/\theta_0 + \frac{\partial\overline{p}}{\partial z} = \overline{\mathbf{v}}.\nabla\overline{w} - \overline{\mathbf{v}.\nabla w}$$

$$\nabla.\overline{\mathbf{v}} = 0 \qquad (80)$$

$$\frac{\overline{D}\overline{\theta}}{Dt} = \overline{\mathbf{v}}.\nabla\overline{\theta} - \overline{\mathbf{v}.\nabla\theta}$$

$$\frac{\overline{D}\overline{r}}{Dt} = \overline{\mathbf{v}}.\nabla\overline{r} - \overline{\mathbf{v}.\nabla r}$$

$$\overline{\mathbf{v}}.\mathbf{n} = 0 \text{ on } \partial\Gamma$$

$$\frac{\overline{D}}{Dt} \equiv \frac{\partial}{\partial t} + \overline{\mathbf{v}}.\nabla$$

The overbar denotes the averaging operator, which can be any form of low pass filter is space and time. The sub-grid correlation terms are written in exact form on the right hand sides. It can be considered as a bilinear operator acting on the prognostic quantities $(\overline{\mathbf{v}}, \overline{\theta}, \overline{r})$. Gent and McWilliams propose a specific way of representing the subgrid correlations in terms of the averaged quantities. They first split the operator into antisymmetric and symmetric parts. The antisymmetric part applied to a single variable ϕ can always be written as a pseudo-advection $\mathbf{V}.\nabla\overline{\phi}$. If the only nonlinearity in the equations were a set of terms $\mathbf{v}.\nabla\overline{\phi}_n$, for a set of scalars ϕ_n, then the antisymmetric part of the subgrid model could be written as $\mathbf{V}.\nabla\overline{\phi}_n$ with the same pseudo-advection velocity for each n. As an example of this, consider the case of two-dimensional motion in hydrostatic balance. Assuming there is no variation in the y direction, the y momentum and thermodynmaic equations from (80) become

$$\frac{\overline{D(v+fx)}}{Dt} = \overline{\mathbf{v}}.\nabla\overline{(v+fx)} - \overline{\mathbf{v}}.\nabla(v+fx)$$

$$\frac{\overline{D}\overline{\theta}}{Dt} = \overline{\mathbf{v}}.\nabla\overline{\theta} - \overline{\mathbf{v}.\nabla\theta} \qquad (81)$$

$$\frac{\overline{D}\overline{r}}{Dt} = \overline{\mathbf{v}}.\nabla\overline{r} - \overline{\mathbf{v}.\nabla r}$$

Gent and McWilliams formally do this for general three-dimensional flow, by adding or subtracting appropriate terms to each side of the momentum equations. They thus write (80), with the hydrostatic approximation, as

$$\frac{D^*\overline{\mathbf{v}}_h}{Dt} + (-fV, fU) + \nabla_h\overline{p} = \nabla.\mathbf{E}_h$$

$$g\overline{\theta}/\theta_0 + \frac{\partial\overline{p}}{\partial z} = 0$$

$$\nabla.\mathbf{V} = 0 \qquad (82)$$

$$\frac{D^*\overline{\theta}}{Dt} = 0$$

$$\frac{D^*\overline{r}}{Dt} = F_r$$

$$\mathbf{V}.\mathbf{n} = 0 \text{ on } \partial\Gamma$$

They write out the full forms of E_h and F_r in their paper. Here, D^*/Dt represents advection by a velocity $V = \overline{\mathbf{v}} + \mathbf{v}'$, where $\mathbf{v}'$ is the 'subgrid velocity', which they choose

an explicit formula for in their paper. These equations are identical to (80), apart from the use of hydrostatic balance. The usefulness of the form (82) depends on being able to model E in terms of known averaged quantities, which will be easier if it can be argued that E is small. The right hand side of the θ equation is set to zero because of their assumption that there are no subgrid fluxes across isentropic surfaces (isopycnal surfaces in the ocean). The use of the 'total' velocity $\mathbf{V}$ in the Coriolis term and the choice that $\mathbf{V}$ satisfies the continuity equation and boundary conditions means that the right hand side of the momentum equation in (82) takes the form of a divergence as shown, and means that non-acceleration theorems are respected.

We can now make a link between (82) and reduced systems of equations such as (56). If E_h is either zero, or can be represented explicitly in terms of the mean quantities, and the averaging scale is sufficiently large for $\overline{\mathbf{v}}_h$ and $\overline{\theta}$ to be in geostrophic and hydrostatic balance, then (82) and (56) become identical, with a forcing term added to the right hand side of the momentum equation in (56). The existence theory for (56) shows that $\mathbf{V}$ is uniquely determined. Thus the part of the subgrid model represented by the pseudo-advection plays the role of maintaining large scale balance and is completely determined by the large scale balance requirement. In addition, the equations can be solved with $E = 0$, in which case exactly the semi-geostrophic solution is obtained. Since this solution can be discontinuous, it implies the averaging operator must be a parcel average, rather than an Eulerian average.

If this type of formulation is used for the moist convectively unstable case discussed in the previous subsection, then $\mathbf{V}$ becomes a generalised flow, and its effect cannot be represented as simple advection, but as an antisymmetric operator acting on a full vertical column of values of temperature and moisture.

The link between the specific reduced system (56) and (82) is natural because both are likely to be most accurate when the flow is approximately two-dimensional. A similar identification should be possible for systems describing approximately axisymmetric flows. More generally, Hamiltonian reduced systems which represent the fluid trajectory separately from the momentum, such as some of those described in the later chapters by Roulstone and McIntyre and by Holm in this volume, can be given a physical interpretation in terms of a mean velocity and a total trajectory that includes subgrid transport.

The practical implications of this link in designing subgrid models for operational atmospheric models are discussed in section 7. The physical implications are also important. For instance, a fundamental part of the general circulation of the atmosphere is the average ascending motion in the tropics. Much of this represents the overall effect of a large number of individual convective events, on quite small scales. Emanuel et al. (1994,p.1125) state that 'nearly all the upward motion associated with ensemble averaged ascent must appear as increased mass flux in cumulus clouds'. In this formulation, the increased upward motion would appear as an increased generalised flow $\mathbf{V}$.

5.5 Treatment of unstable regions

The previous sections have emphasised the importance of understanding how the subgrid model interacts with flows that can be described by a reduced set of equations. As discussed in section 2, these equations can typically only describe stable states of the

atmosphere. Observations indicate that, on average, it is possible to have long-lined unstable regions where there is strong forcing. The best example is the near-surface layers of the atmosphere over deserts strongly heated by the sun. This situation can also occur where air is being cooled by precipitation at a rate too fast for the air to be assumed to remain in a stable position. This can happen in large scale convective systems. Shutts et al. (1988) include this effect in a reduced system of equations, but the requirement of large scale balance is probably unreasonable, and the resulting solution may exaggerate the effect of the precipitation.

It is simplest to consider the first example. In (80), assume that $\frac{\overline{\partial\theta}}{\partial z}$ is negative, then the mean fields are statically unstable and N^2 as given by (9) is negative. The left hand side of (80) has solutions which grow exponentially. Assuming a two-dimensional wavelike perturbation $\exp(\omega t + i(kx + mz))$, the growth rate is given by

$$\omega = \frac{k}{m}\sqrt{(-N^2)} \tag{83}$$

The growth rate is largest on the smallest horizontal scale permitted by the averaging, and the largest vertical scale permitted by the atmosphere. A subgrid model which is correct in the sense of allowing quasi-steady solutions to persist has to exclude these growing solutions, or reduce the growth rate to be comparable to observed timescales of variability. For example, suppose we choose a subgrid model which diffuses in the x direction, where the maximum growth rate is on the smallest scale. A linearised version of (80) with such a model included is then

$$\begin{aligned}
\frac{\partial\overline{u}}{\partial t} + \frac{\partial\overline{p}}{\partial x} &= K\frac{\partial^2\overline{u}}{\partial x} \\
g\overline{\theta}/\theta_0 + \frac{\partial\overline{p}}{\partial z} &= 0 \\
\frac{\partial u}{\partial x} + \frac{\partial w}{\partial z} &= 0 \qquad (84) \\
\frac{\partial\overline{\theta}}{\partial t} + w\frac{\partial\overline{\theta}}{\partial z} &= K\frac{\partial^2\overline{\theta}}{\partial x^2} \qquad (85)
\end{aligned}$$

Assuming solutions proportional to $\exp(\omega t + i(kx + mz))$ gives

$$(\omega^2 + k^2K)^2m^2 + k^2N^2 = 0 \tag{86}$$

This has no growing solutions if $K > (\sqrt{(-N^2)})/mk$. If K is chosen to be greater than this value where m and k are the smallest wavenumbers permitted by the domain, or by the region over which N^2 is negative, the sub-grid model will be consistent with the persistence of the observed unstable region. The values of diffusion required will typically be large, since it will have to damp on a timescale comparable to the growth rate of the instabilities multiplied by k/m. Following (83), the largest value of k/m permitted by the size of the unstable region corresponds to the maximum growth rate, and must therefore be used in calculating the damping timescale required. This diffusion should only be used in unstable regions of the flow.

5.6 Example of modelling of the response to thermal forcing by a reduced set of equations

In this subsection we give an example of how the large scale response to atmospheric forcing can be described by the solution of a reduced set of equations with the appropriate forcing terms included. We thus solve (77) on the sphere, with a given heat source in the Northern hemisphere subtropics and a mountain ridge crossing the equator confining the response to the forcing. The experiment is described by Cullen and Mawson (1992).

The last equation in (77) shows that, outside the boundary layer, $\nabla_h s$ must be small near the equator. Differentiating with respect to z shows that $\nabla_h \theta$ will be small. The total velocity $\mathbf{u}$ can be found from (64) with (78) and (79). If the heating function S_h is localised, a steady solution for $\mathbf{u}$ will be approximately given by

$$w \frac{\partial \Theta}{\partial z} = S_h \tag{87}$$

This suggests that there will be ascending motion in the region of the heat source, with inflow from other regions at low levels to balance it. (78) shows that the largest response in the horizontal wind will be in directions where $f^2 + c_D^2$ is smallest, Schubert et al (1991), so that the greatest response will be from the equatorial side of the heat source above the boundary layer. In the boundary layer, (74) shows that a pressure gradient will be set up parallel to the inflow.

The experiment includes a mountain ridge crossing the equator to the west of the region where the heat source is applied. This should have the effect of confining the inflow towards the heated region, and locally inducing a maximum cross-equatorial wind on the eastern side of the ridge.

The results are illustrated in Fig. 11. This shows the low level wind induced by the heat source. The coastlines are included for illustrative purposes only. The heat source is centred over Northern India and the mountain ridge croses the equator in east Africa. The solutions of (77) show a maximum cross-equatorial flow on the eastern side of the ridge, and a circulation around the heat source. Solutions of a similar model based directly on (71) are also shown. It is clear that the bulk of the solution is represented by the reduced equations (77), though there are detailed differences. Observations confirm the existence of a low level jet in the area shown.

6 Evolutionary properties of atmosphere/ocean circulations

6.1 General remarks

In this section we seek to derive a picture of the large scale dynamics of the atmosphere or ocean as a dynamical system. The basic circulation is symmetric about the poles, with higher temperatures at the equator, as a result of the pole/equator radiation difference. There are permanent asymmetries in the circulation, which are most likely to be a response to the asymmetric land-sea and mountain distribution, particularly in the Northern hemisphere. The Southern hemispheric circulation is much more symmetric. The total circulation also has large transient asymmetries. It is permanently unsteady, but with occasional episodes when there are large quasi-steady circulations. These are called blocking patterns, and can give anomalous weather over periods from weeks to a

few months. Qualitatively, the weather map looks much the same from day to day (Fig. 12). The characteristic scale of weather systems are always essentially the same. Deterministic prediction is, however, limited to a few life-cycles of the individual weather systems.

It is natural, therefore, to seek a model of the underlying large scale dynamics which has stationary statistics and gives quasi-periodic solutions which can be continued indefinitely in time. The asymmetries and seasonal variations would then follow when the boundary conditions and forcing terms were included. Existing theory does not achieve this, because no argument has been found which prevents a systematic geeration of small scales in the flow. The stationary statistics of the observed flow have to be explained by a balance between forcing and dissipation.

Since the pioneering work of Lorenz (1963), there have been many attempts to describe the qualitative features of the large scale dynamics discussed above using low order systems of equations. In this section we, instead, exploit the property, discussed in section 4, that the solutions to the general equation stay close to those of a reduced system. As discussed in section 4.5, we use a reduced system which is Hamiltonian, so in particular has a conserved energy and potential vorticity. The Hamiltonian property means there can be no attractors.

The general method is to exploit the simultaneous conservation of energy, potential vorticity and potential temperature. We first show that the condition for the flow to be steady is that it is a stationary point of the energy with respect to perturbations which simultaneously rearrange the potential vorticity and potential temperature. This type of condition is a general characterisation of steady states of Hamiltonian systems, see Abarbanel and Holm (1987). We then use the principle of Kelvin (Thomson (1910)) to say that stable steady states correspond to extrema of the energy. A typical result would be that there are only two globally stable steady states associated with a given potential vorticity and potential temperature distribution , one achieved by maximising and one achieved by minimising the energy. The actual result in a particular case will depend on the boundary conditions and physical restrictions on the permitted rearrangements.

Given an initial distribution of potential vorticity and potential temperature, construct the minimum energy state. If the initial data has greater energy, the evolution will then remain a fixed distance in energy above this basic state. A Hamiltonian system has no attractors. Thus if the initial data represents an unsteady state, it cannot evolve to a steady state, nor can it be attracted to a periodic solution as this would be an attractor. If it is a steady state, it cannot be guaranteed to be nonlinearly stable unless the energy is equal to the minimising or maximising value.

More information can be obtained by seeking locally stable states, which are extrema of the energy with respect to continuous displacements, a subset of the global rearrangements. This can be used to explain the long-lived anomalies in atmospheric or oceanic flows. However, a general solution cannot be attracted to the neighbourhood of such a steady state, as that would make it an attractor. Non-Hamiltonian forcing terms must be included to make this happen.

The other results we discuss concern the preservation of the statistics of the flow. As shown in section 4.5, the reduced systems of equations commonly used to describe

large scale atmospheric motions have a structure similar to the equations governing two-dimensional incompressible flow. The solutions of the two-dimensional Euler equations describe 'two dimensional turbulence', Leith (1983), with a k^{-3} energy spectrum. Theory and computation indicates that such flows exhibit a migration of the energy to the largest available scale, and the enstrophy to the smallest scales, the 'enstrophy cascade'. It then has to be argued that the observed preservation of the mean scale of disturbances is due to the systematic injection of energy at particular scales by non-conservative effects.

We discuss this behaviour first in terms of bounds on the L_2 norm of the velocity and its gradient. If both these are separately conserved, or bounded uniformly in time, then the mean scale is also bounded in time. (This argument was first advanced by Fjortoft (1953)) based on energy and enstrophy conservation). We also discuss the issue in terms of local estimates on the velocity gradients. Control of these would prevent systematic generation of small scales even in small regions of the flow, and thus give greater control over the flow statistics.

6.2 Stable steady states

Many of the results available can be simply illustrated by the equations for two-dimensional incompressible flow. These can be written on a region Γ as

$$\begin{aligned} \frac{D\zeta}{Dt} &= 0 \\ \frac{D}{Dt} &\equiv \frac{\partial}{\partial t} + \mathbf{U}.\nabla\zeta \\ \mathbf{U} &= (-\frac{\partial\psi}{\partial y}, \frac{\partial\psi}{\partial x}) \\ \zeta &= \nabla^2\psi \\ \mathbf{U}.\mathbf{n} &= 0 \text{ on } \delta\Gamma \end{aligned} \tag{88}$$

Burton and McLeod (1991) proved that if the region Γ is a disk, the energy is maximised by a symmetrical arrangement of vorticity ζ monotonically increasing with streamfunction, and minimised by a symmetrical arrangement of vorticity ζ monotonically decreasing with streamfunction. These results and others are described in the chapter by Burton and Douglas later in this volume.

Applications of this type of theory to large scale atmospheric flow requires inclusion of the variation with latitude of the vorticity associated with the Earth's rotation. The equations for two-dimensional incompressible flow on a rotating spherical surface are

$$\begin{aligned} \frac{D}{Dt}(\zeta + f) &= 0 \\ \frac{D}{Dt} &\equiv \frac{\partial}{\partial t} + \mathbf{U}.\nabla\zeta \\ \mathbf{U} &= (-\frac{\partial\psi}{\partial y}, \frac{\partial\psi}{\partial x}) \\ \zeta &= \nabla^2\psi \\ f &= 2\Omega\sin\phi \\ \mathbf{U}.\mathbf{n} &= 0 \text{ on } \partial\Gamma \end{aligned} \tag{89}$$

Solutions of these equations rearrange of the initial absolute vorticity $\zeta + f$. The energy is conserved, but depends on the streamfunction which is calculated only from the relative vorticity ζ. The angular momentum is conserved. The planetary vorticity f is given by $2\Omega\cos\phi$, which is a smooth monotonic function of latitude. The natural conjecture is that the rearrangements of absolute vorticity which extremise the energy are functions of latitude only, and are monotonically increasing or decreasing with latitude. Only the minimum energy state, corresponding to the absolute vorticity increasing with latitude, as f does, is likely to be physically realisable.

A difficulty in this case will be the possibility of 'mixing' rearrangements, as discussed in section 3.7. If the given distribution of $\zeta + f$ can be mixed, subject to angular momentum conservation, to give a distribution equal to f, the minimum energy reachable will be zero and there will be no non-trivial stable steady state. It is clear that this cannot happen if the given values of $\zeta + f$ do not cover the full range of values of f. Thus there will be stable steady states with negative relative vorticity at the North pole, and positive at the South pole. This corresponds to a basic easterly flow, which is opposite to that required to balance the equator-pole temperature difference. The stability of westerly basic flows against this type of mixing has still to be examined.

A more complete reduced model is the semi-geostrophic system (56). This can be written as an evolution equation for an inverse potential vorticity in the form (60), (61). We first illustrate stability results for it where the solution region Γ is a channel of width $2D$ and height H, with periodicity $2L$ in the x direction. The evolution equations (60) take the form

$$\begin{aligned} \frac{D\rho}{Dt} &= 0 \\ \frac{D}{Dt} &\equiv \tfrac{\partial}{\partial t} + U\tfrac{\partial}{\partial X} + V\tfrac{\partial}{\partial Y} + W\tfrac{\partial}{\partial Z} \\ U &= f(\tfrac{\partial R}{\partial Y} - Y) \\ V &= X - f(\tfrac{\partial R}{\partial X}) \\ W &= 0 \end{aligned} \tag{90}$$

R is determined from the Monge Ampere equation (34), with $\nabla R = (x, y, z)$, and the boundary conditions (61) become

$$\begin{aligned} \frac{\partial R}{\partial X}(X,Y,Z) &= \tfrac{\partial R}{\partial X}(X+2L,Y,Z) \\ -D \le \frac{\partial R}{\partial Y} &\le D \\ 0 \le \frac{\partial R}{\partial Y} &\le H \end{aligned} \tag{91}$$

It is convenient to define

$$\Psi = R - \frac{1}{2}(X^2 + Y^2) \tag{92}$$

Ψ acts as a streamfunction for the flow defined in (90). The problem is to be solved in (X, Y, Z) space $(-L, L) \times (-\infty, \infty) \times (-\infty, \infty)$. The compatibility conditions (35) for the Monge-Ampere equation become

$$\rho(X+L,Y) = \rho(X-L,Y) \tag{93}$$
$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-L}^{L}\rho dX dY dZ = 4LDH$$

The class of perturbations which is dynamically consistent with (90) is $\mathcal{R}_h(\rho_0)$, defined by

$$\rho \in \mathcal{R}_h(\rho_0) \text{ if } \begin{cases} \rho(.,Z) \in \mathcal{R}(\rho_0(.,Z)) \text{ for almost all } Z \\ \int Y\rho dX dY dZ = \int Y\rho_0 dX dY dZ \end{cases} \tag{94}$$

The first condition restricts the rearrangements of ρ to the (X,Y) variables only. This is similar to the space of 'stratified' rearrangements used by Burton and Nycander (1996). The additional condition is that the mean Y over the particles cannot be changed. This corresponds to angular momentum conservation which is implied by the periodic boundary conditions, which mean, using (56) that $\int_{-L}^{L} v_b dx = \int_{-L}^{L} \frac{\partial p}{\partial x} = 0$.

We next demonstrate the characterisation of steady states of the semi-geostrophic system in terms of stationary points of the energy with respect to rearrangements following the principle of Abarbanel and Holm (1987). The energy for the problem (90) is given by (44), which can be rewritten in the form (36) as

$$\begin{aligned} E = & \int_{-H}^{H}\int_{-D}^{D}\int_{-L}^{L}\{\tfrac{1}{2}f^2\left((X-x)^2+(Y-y)^2\right) - zZ\}dxdydz \\ = & \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-L}^{L}\{\tfrac{1}{2}f^2\left((X-x)^2+(Y-y)^2\right) - zZ\}\rho dX dY dZ \end{aligned} \tag{95}$$

Given $\rho = \rho_0(X,Y,Z)$ satisfying (67), seek the conditions under which $\delta E = 0$ for perturbations to ρ satisfying $\rho + \delta\rho \in \mathcal{R}(\rho_0)$. These can be generated by keeping ρ fixed on particles in $\mathbf{X}$ space and perturbing X and Y with a displacement field χ. The displacement must be non-divergent, so can be written as $(-\frac{\partial\alpha}{\partial Y}, \frac{\partial\alpha}{\partial X}, 0)$ for an arbitrary function $\alpha(X,Y,Z)$, and must satisfy the periodicity condition so that $\alpha(X-L,Y,Z) = \alpha(X+L,Y,Z)$. The restriction that the mean Y cannot be changed is automatically enforced by the periodicity condition. We then have

$$\delta E = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-L}^{L}((X-x)\delta X + (Y-y)\delta Y - X\delta x - Y\delta y - Z\delta z)\rho dX dY dZ \tag{96}$$

where the integration is taken over particles, so that there is no $\delta\rho$, and we have used the invariance of $\int_0^H\int_{-D}^{D}\int_{-L}^{L} x^2 + y^2 dxdydz$. The energy minimisation property of the solutions of (90) discussed in section 3.5 means that

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-L}^{L}(-X\delta x - Y\delta y - Z\delta z)\rho dX dY dZ = 0 \tag{97}$$

because this is the variation of $\mathbf{x}$ with $\mathbf{X}$ fixed used in (27). Substituting the definitions of Ψ and $\delta\mathbf{X}$ gives after some manipulations

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-L}^{L}\left(\nabla.\left(\alpha\rho\frac{\partial\Psi}{\partial Y},-\alpha\rho\frac{\partial\Psi}{\partial X}\right)-\alpha\left(-\frac{\partial\Psi}{\partial X}\frac{\partial\rho}{\partial Y}+\frac{\partial\Psi}{\partial Y}\frac{\partial\rho}{\partial X}\right)\right)dXdYdZ=0 \quad (98)$$

Assuming that ρ vanishes at a sufficiently large $|Y|$ and requiring (98) to hold for arbitrary α gives

$$-\frac{\partial\Psi}{\partial X}\frac{\partial\rho}{\partial Y}+\frac{\partial\Psi}{\partial Y}\frac{\partial\rho}{\partial X}=0 \quad (99)$$

This condition is precisely that for the flow to be steady as we can see from (90) and (92).

Following Kelvin's principle, we next characterise steady states which are stable by requiring the stationary point of the energy to be an extremum. We construct a series of examples to illustrate the results of doing this. As shown by Burton and Nycander (1996), any $\rho(X,Y,Z)$ satisfying (93) with compact support for each Z, so that ρ is zero for $|Y|>Y_0$ for some $Y_0(Z)$, can be rearranged into a distribution corresponding to a steady shear flow parallel to the X axis with ρ increasing in Y:

$$\begin{aligned}\rho_0(X,Y,Z) &= g_0(Y,Z)\\ \int_{-\infty}^{\infty}g_0(Y,Z)dY &= 2D\ \forall Z \quad (100)\\ \frac{dg_0(Y)}{dY} &\geq 0\end{aligned}$$

We then seek to extremise E for $\rho\in\mathcal{R}(\rho_0)$. It is first clear that the maximum energy attainable under these conditions is infinite. Generate a rearrangement by a displacement $\alpha=A\sin(\pi kX/L)$, implying $\delta X=0,\delta Y=\frac{A\pi k}{L}\cos\left(\frac{\pi kX}{L}\right)$. Since $|y|<D$ for all particles, (95) shows that $E\to\infty$ as $A\to\infty$. It is therefore only meaningful to seek minimum energy states. Note that the maximum energy state was well defined for the barotropic vorticity equation (88), because the displacements χ of vorticity have to be within the domain, so that $\alpha=0$ on the domain boundaries.

In general, the minimum energy is attained by making the map from (X,Y,Z) to (x,y,z) as close as possible to the identity map $x=X,y=Y,z=Z$ which has zero energy. Consider some specific examples. First consider the sequence of Z independent distributions

$$\rho_n(X,Y,Z)=\begin{cases}2f_n(\frac{X+L}{2L}) & \text{if } -D\leq Y\leq D, 0\leq Z\leq H\\ 0 & \text{otherwise}\end{cases} \quad (101)$$

where f_n is as defined in (41) and (42). The solution $\mathbf{x}=\mathbf{x}(\mathbf{X})$ is required to satisfy

$$\begin{aligned}\frac{\partial y}{\partial Y} &= 1\\ \frac{\partial z}{\partial Z} &= 1\\ \frac{\partial x}{\partial X} &= \begin{cases}0 & \text{if } X=-L,\\ 0 & \text{if } X\in((2m-n)L/n,(2m+1-n)L/n],\\ 2 & \text{if } X\in[(2m+1-n)L/n,L(2m+2-n)/n],\end{cases}\end{aligned}$$

where $m = 0, 1, ..., n-1$.

The energy associated with this solution is $2HL^3D/3n^2$. As n tends to infinity, the solution tends weakly to $\rho = 1 : -D \le Y \le D, 0 \le Z \le H$ so that $\mathbf{X} = \mathbf{x}$. The energy in the limit case is zero.

In this example, the minimum energy is attained by making ρ as close to 1 as possible. This is because $\rho = 1$ corresponds to the identity map $x = X, y = Y, z = Z$ which has zero energy. It may be possible to achieve it by mixing values as in section 3.7 to obtain a mean of 1. In general, we expect that the only candidates for minimum energy states among distributions of the form (100) will be those with $g_0(Y) \le 1$ everywhere, because no mixing can give an increased range of values of Y for which $g(Y) = 1$. Since ρ and g are inverse potential vorticities, this condition excludes values of potential vorticity less than 1, which correspond to anticyclonic relative vorticity. This agrees with the result of Kushner and Shepherd (1995) that there were no stable shear flows with anticyclonic shear. This argument would also show that no steady states with anticyclonic relative vorticity were stable in a limited domain under semi-geostrophic dynamics. They could be stable in doubly periodic flows, because there is then no region of $\rho = 0$ in the (X, Y) planes to mix with the non-zero values.

Note also that the Z independence assumed in (101) is needed to allow a zero energy to be reached. If the given distribution depends on Z, it may be possible to construct a sequence of rearrangements along the lines of (101) which converges weakly to a distribution independent of Z. However, the z and Z values enter the energy (95) as a term $-zZ$, not as $(z - Z)^2$. This means that the minimum energy will be given by an arrangement with z as strongly correlated with Z as possible, and a lower energy cannot be reached by mixing. It is therefore likely that the minimising energy corresponds to a separate minimisation on each Z surface, without restricting the mean Y, followed by shifting the mean Y an equal amount for all Z to satisfy the angular momentum constraint. Fig. 13 shows the distribution of ρ corresponding to a baroclinic shear flow with U increasing with z. The minimum energy rearrangement is a flow with no vertical shear and a velocity $\overline{U}(z)$.

Noting this, we now impose the condition $g_0(Y) \le 1$ in (100), and minimise the energy with respect to distributions of ρ of the form $\rho(X, Y, Z) = g(Y)$, where $g \in \mathcal{R}(g_0)$. The energy density given by this distribution is

$$2LH \int_{-\infty}^{\infty} \frac{1}{2} \left(\int_{-\infty}^{Y} g(Y')dY' - Y \right)^2 g(Y)dY \tag{102}$$

$\int Y g(Y)$ is fixed by (94) and $\int g(Y)dY$ is fixed by (100). We give the results of an explicit calculation for the three cases

$$g_0(Y)/4D = 0 \text{ except } \begin{cases} 1-\eta & \text{where } \eta = Y - 1/6, 0 \le \eta \le 1 \\ \eta & \text{where } \eta = Y + 1/6, 0 \le \eta \le 1 \\ 2Y & \text{if } 0 \le Y \le 1/2,\ 2 - 2Y \text{ if } 1/2 \le Y \le 1. \end{cases} \tag{103}$$

These are illustrated in Fig.14. The different definitions of η are chosen so that the three distributions have the same mean Y, in order to satisfy the constraint (94). The first two represent respectively monotonically decreasing and increasing profiles of ρ. The overall symmetry of the problem means that these will have equal energy. The third is a symmetrical distribution of ρ, but has a vanishing vorticity gradient. It would be expected

from the Charney-Stern theorem that the third state would not be stable, and so would not be the energy minimiser. Explicit calculation confirms this by showing that the energy associated with the three, setting $y = y_0$ at $Y = -\infty$ in each case, is $y_0^2/4 - y_0/8 + 13/720$ for the first two choices, and $y_0^2/4 - y_0/8 + 11/192$ for the third. The first two represent stable steady states, as no lower energy can be reached. This is in agreement with the results of Kushner and Shepherd, but is a stronger result than theirs which was restricted to small amplitudes.

This suggests that, in general, stable steady flows in the barotropic semi-geostrophic case correspond to monotonically increasing or decreasing ρ distributions which are independent of X and have $\rho \leq 1$. We now need to check that the result survives mixing of the profile. As in section 3.7, we can construct a sequence of rearrangements of the profile $g_0(Y)$ defined in (100) that converges to a uniform value of g determined by (94). In the example (100), the value would be 1/2. The energy associated with this is $y_0^2/4 - y_0/8 + 1/48$. This is greater than the value for the monotonic profiles of (100), suggesting that these two states retain their stability even when all weak limits of rearrangements are considered. It remains to be proved under what circumstances the same holds for general profiles of ρ other than the linear one assumed in (100).

In the three-dimensional semi-geostrophic case, the stable steady states will be monotonic rearrangements of ρ on each Z surface, aligned so that the Z variations of the ρ distribution is minimised. Burton and Nycander (1996) prove a similar result for the three-dimensional quasi-geostrophic case. In the first example distribution of (103), the minimisation is obtained by a shift so that $\eta = Y - 1/12$. Thus if ρ for each Z belongs to a single rearrangement class $\mathcal{R}(\rho_0)$, the stable steady state is barotropic. A stable state with baroclinic shear is only possible when this is not the case, for example, if ρ is given as either 0 or 1 everywhere, but the measure of the region with non-zero ρ varies with Z, (Fig. 13(c)). The latter situation certainly applies when considering the stability of the mean circulation of the global atmosphere, and means that zonally symmetric states with monotonic variation of ρ on each isentropic surface can be stable. These conditions are sufficient conditions for stability. They appear more restrictive than the necessary conditions resulting from linear baroclinic instability theory, but this requires further exploration.

6.3 Locally stable states

The global stability results discussed above require extremisation of the energy with respect to any rearrangement of the potential vorticity, even if the rearrangement cannot be generated by any possible velocity field. Burton and Nycander (1996) showed that a local patch of anomalous potential vorticity in a background flow with uniform potential vorticity can be nonlinearly stable as a maximum energy state using the quasi-geostrophic equations with respect to distortions of the patch. This result is described in the chapter by Burton later in this volume.

Blocking patterns in the atmosphere have a horizontal scale comparable to the Earth's radius. Explaining them as local stable steady states requires inclusion of the variation with latitude of the vorticity associated with the Earth's rotation. Equations such as (89) which include this support dispersive Rossby waves, which makes it difficult to maintain a steady localised perturbation to a zonally symmetric flow. The simplest type of solution

for a finite amplitude steady state is a Rossby wave superposed on a mean flow which cancels out the wave speed. A solution of (89) with modified boundary conditions which illustrates this is as follows:

$$\begin{aligned} \Gamma &= (-L, L) \times (0, D) \\ f &= \beta y \\ \zeta &= A \sin\left(\frac{\pi k x}{L}\right) \sin\left(\frac{\pi l y}{D}\right) \qquad (104) \\ \psi &= -\frac{A}{\pi^2 k^2/L^2 + \pi^2 l^2/D^2} \sin\left(\frac{\pi k x}{L}\right) \sin\left(\frac{\pi l y}{D}\right) - \frac{\beta y}{\pi^2 k^2/L^2 + \pi^2 l^2/D^2} \end{aligned}$$

This is a steady solution for any given k and l. It satisfies the boundary conditions that ψ and $\mathbf{u}$ are periodic in x with period $2L$, and $v = 0$ on $y = 0, D$. The dispersive nature of Rossby waves means that this type of solution can only be obtained for monochromatic waves. This has been exploited by Neven (1994) who constructed a set of localised 'modon' steady solutions by patching together monochromatic solutions.

Another example is a solution of the semi-geostrophic system (90) generalised to include variable Coriolis parameter using the method of section 3.6. The condition for a steady state is as before

$$\Psi = \Psi(\rho) \qquad (105)$$

where Ψ is the streamfunction for the flow but now $\rho = \frac{1}{f}\frac{\partial(x,y)}{\partial(X,Y)}$. The only solution found for this so far is an annular region or patch of zero absolute vorticity (Fig.15), so that a finite region in $\mathbf{x}$ space maps onto a single circuit or point $\mathbf{X}_0$ in $\mathbf{X}$ space and ρ is a distribution. The boundaries of the annular region must satisfy the condition

$$\mu(\mathbf{x}, \mathbf{X}_0) = \text{constant} \qquad (106)$$

where μ is defined by (39). This solution would correspond to a localised high pressure centre, with a central region of zero pressure gradient surrounded by a region of strong anticyclonic shear.

6.4 Regularity estimates and the enstrophy cascade

The generation of small scales in the enstrophy in two dimensional incompressible flow governed by (88) or (89) can be understood using the analytical tools which are used to prove existence and regularity of these equations. The rigorous theory we draw on is set out in Gerard (1992). The equations state that the absolute vorticity $\zeta + f$ is a rearrangement of the initial distribution. Existence proofs by the standard methods require that the solutions are contained in a compact set. However, as discussed in section 3.7, the set of rearrangements is not compact. We therefore have to identify a "reachable" compact subset of the set of rearrangements - rearrangements that can actually be reached in the time evolution. Given initial absolute vorticity which has bounded gradients, we can do this by estimating the rate of growth of the vorticity gradients. (89) implies that

$$\frac{D}{Dt}\nabla(\zeta + f) + \begin{pmatrix} \frac{\partial u}{\partial x} & \frac{\partial v}{\partial x} \\ \frac{\partial u}{\partial y} & \frac{\partial v}{\partial y} \end{pmatrix} \nabla(\zeta + f) = 0 \qquad (107)$$

This can be written in terms of the streamfunction ψ as

$$\frac{D}{Dt}\nabla(\zeta+f)+\begin{pmatrix} -\frac{\partial^2\psi}{\partial x\partial y} & \frac{\partial^2\psi}{\partial x^2} \\ -\frac{\partial^2\psi}{\partial y^2} & \frac{\partial^2\psi}{\partial y\partial x}\end{pmatrix}\nabla(\zeta+f)=0 \tag{108}$$

These equations can be used to estimate the rate of increase of vorticity gradients, using a bound on the velocity gradients in terms of the vorticity and its gradients (Gerard (1992),p 424):

$$\| \nabla\mathbf{U} \| \leq C \| (\zeta+f) \| \log\left(1+\frac{\| \nabla(\zeta+f) \|}{\| (\zeta+f) \|}\right) \tag{109}$$

Exact definitions of the norms used are given in Gerard. However, they are essentially maximum norms. The bound is derived from the solution procedure for the Poisson equation for ψ in terms of ζ. Because of the dependence of the bound in (109) on the vorticity gradients, the estimate of vorticity gradients obtained from (107) allows exponential growth in time. This does not prevent regularity being proved for all time, but allows the accumulation of enstrophy at small scales. This can be expressed as a statement that

$$\| \nabla\mathbf{u}^2 \| \leq C \| \mathbf{u}^2 \| \tag{110}$$

where C grows exponentially in time. If the L_2 norm is used instead of the maximum norm, then an estimate of the form (110) holds with C independent of time, but dependent on the domain size. Thus the mean scale of the flow is bounded, but local regions where small scales are generated are permitted. This agrees with widesporead computational experience.

Now consider the semi-geostrophic model, as expressed by (60) and (61). An equation of the form (108) still governs the rate of growth of inverse potential vorticity gradients. Using (92), we obtain

$$\frac{D}{Dt}(\nabla\rho)+\begin{pmatrix} -\frac{\partial^2\Psi}{\partial X\partial Y} & \frac{\partial^2\Psi}{\partial X^2} \\ -\frac{\partial^2\Psi}{\partial Y^2} & \frac{\partial^2\Psi}{\partial Y\partial X}\end{pmatrix}\nabla\rho=0 \tag{111}$$

The equation that determines Ψ from ρ is a Monge-Ampere equation for R, followed by addition of the fixed function $\frac{1}{2}(X^2+Y^2)$ to obtain Ψ. Thus, in effect, ρ is transported by a convex streamfunction. If the flow is steady, convexity prevents exponential growth of gradients of ρ, except for time periods short compared with that needed for a particle to make a complete circuit of a streamline. Only algebraic growth is permitted. For general time dependent solutions of (60) and (61), it is a plausible but unproved conjecture that exponential growth is prevented and a sharper estimate than (109) obtained. If so, there would be no systematic enstrophy cascade in semi-geostrophic turbulence, or in the solutions of more general reduced systems of equations where the Monge-Ampere equation is used to generate other fields from the potential vorticity.

7 Applications to numerical model design

7.1 General considerations

We can summarise the results reviewed in the previous subsections that are important for the design of practical atmospheric numerical models as follows. Most of the points apply equally to ocean models.

i) In a given asymptotic regime, the solution of the full equations stays close to that of an appropriate reduced system (section 4).

ii) Most useful simple models appropriate for the weather and climate forecasting problem can be described in terms of advecting a conserved potential vorticity, from which the other fields can be derived (section 4.5).

iii) An accurate representation of potential vorticity conservation is therefore needed to give accurate predictions of the adiabatic evolution of the atmosphere.

iv) The solution procedure for the full equations needs to respect the link between the potential vorticity and the other fields in the appropriate reduced system of equations.

v) The subgrid model forms part of the method by which the overall solution stays close to that of the reduced system. (section 5.4)

vi) The solution procedure should control the distance by which the solution departs from that of the reduced system to the theoretically predicted level.

7.2 Semi-implicit methods and relation to schemes for reduced equations

We illustrate these points by using (71) as the model, thus including the sub-grid model, but avoiding other complications present in (1). A standard algorithm for solving these equations, and for solving (1), uses 'semi-Lagrangian' methods for advection terms and an implicit scheme for calculating the pressure, Staniforth and Cote (1991). A simple form of such a scheme can be written as

$$
\begin{aligned}
\mathbf{u}_h^{n+1} &= \mathbf{u}_{hd}^n - \delta t \left((-fv, fu)^{n+1} + \nabla_h p^{n+1} - (F_u, F_v)^{n+1} \right) \\
w^{n+1} &= w_d^n + \delta t \left(g\theta^{n+1}/\theta_0 - \frac{\partial p^{n+1}}{\partial z} \right) \\
\nabla.\mathbf{v}^{n+1} &= 0 \qquad (112) \\
\theta^{n+1} &= \theta_d^n + \delta t \left(F_h^{n+1} + H - LP + S_h \right) \\
r^{n+1} &= r_d^n + \delta t \left(F_r^{n+1} + P + S_r \right) \\
\mathbf{v}^{n+1}.\mathbf{n} &= 0 \text{ on } \partial\Gamma
\end{aligned}
$$

Superscripts n and $n+1$ denote discrete time levels. δt is the timestep. The calculation is assumed to be performed at discrete gridpoints. Suffix d denotes a value calculated at the 'departure point' of a trajectory which finishes at the gridpoint being calculated at

the end of the timestep. Values at departure points have to be obtained by interpolation from the nearest gridpoints. In practice second order time accuracy is used, rather than the first order scheme written out in (112).

In (112) the sub-grid terms $(F_u, F_v, F_\theta, F_r)$ are all computed at time level $n+1$. This is because the momentum terms (F_u, F_v) form part of the lowest order balance as discussed in section 5.2. The time integration is designed so that the lowest order balance will be satisfied at time level $n+1$ if the acceleration terms are zero. It was shown in section 5.5 that the subgrid term in F_θ has to be considered along with the explicit dynamics when showing that the averaged evolution of statically unstable regions has a slow timescale. If the model is to be stable with a timestep appropriate for the slow timescale, but long compared with the instability that would be generated by the explicitly resolved dynamics, the term F_θ has be evaluated at time level $n+1$. A similar argument for moist statically unstable regions shows that the term F_r also has be evaluated at time level $n+1$.

(71) contains equations for the explicit evolution of all the variables except the pressure. A standard solution procedure is to calculate the pressure at time level $n+1$ by using the condition that the continuity equation and boundary condition are satisfied at time level $n+1$. This is closely analogous to the analytical solution procedure described in section 3.3. It gives

$$\nabla^2 p^{n+1} = \nabla.\mathbf{u}_d^n - \delta t \left(\nabla.(-fv, fu, g\theta/\theta_0)^{n+1} - \nabla.(F_u, F_v, 0)^{n+1} \right) \tag{113}$$

$$\frac{\partial p^{n+1}}{\partial n} = \mathbf{u}_d^n.\mathbf{n} - \delta t \left((-fv, fu, g\theta/\theta_0)^{n+1}.\mathbf{n} - (F_u, F_v, 0)^{n+1}.\mathbf{n} \right) \text{ on } \partial\Gamma$$

In order to solve (113), we have to make a linearised estimate of all the terms on the right hand side evaluated at time level $n+1$. This can be done by using the reduced set of evolution equations (77), which predict the evolution of $\nabla s = (fv_c + c_D u_c, -fu_c + c_D v_c, g\theta/\theta_0)$, precisely the terms that are required apart from F_h. The solution procedure for (77) is based on the form (64), with the matrices $\mathbf{Q}$ and $\mathbf{H}$ given by (78) and (79). Therefore

$$\begin{aligned} (\nabla s)^{n+1} &= (fv_c - c_D u_c, -fu_c - c_D v_c, g\theta/\theta_0)^{n+1} \\ &= (\nabla s)^n + \mathbf{H}^n - \mathbf{Q}\delta t \begin{pmatrix} u \\ v \\ w \end{pmatrix}^{n+1} \end{aligned} \tag{114}$$

In order to obtain a single scalar equation for p^{n+1}, we substitute this into (112) rather than (113). This gives equations for the velocity components at time $n+1$ of the form

$$(\mathbf{I} + \mathbf{Q}\delta t^2)\mathbf{u}^{n+1} = \mathbf{u}_d^n - \delta t \left(\nabla p^{n+1} + (fv, -fu, g\theta/\theta_0)^n + (F_u, F_v, 0)^n \right) \tag{115}$$

This can be reduced to a single equation for p^{n+1} similar to (113) by using the condition $\nabla.\mathbf{u}^{n+1} = 0$.

Successful use of this algorithm depends on the ellipticity of the operator $\mathbf{I} + \mathbf{Q}\delta t^2$. If $\mathbf{Q}$ has no negative eigenvalues, so that the flow is statically and inertially stable, this is assured. Otherwise there is a timestep limit

$$\delta t^2 \leq 1/\lambda \tag{116}$$

where λ is the largest negative eigenvalue of $\mathbf{Q}$. This is a statement that the timestep must be short enough to resolve the physical instabilities that will occur for such data. In operational models, this causes a difficulty, because such regions are small in extent, and it is undesirable to restrict the timestep for the complete model to that required in the most unstable region. Thus the sub-grid model should be designed assuming a time as well as space averaging, and reduce the maximum growth rates of instabilities to be on the timescale required, as shown in section 5.5.

If the timestep is large, and the time level used in the implicit calculations is entirely at time level $n+1$, as illustrated in (115), then as the timestep becomes large, the solution of (115) tends towards

$$\mathbf{Q}\delta t\mathbf{u}^{n+1} = -\left(\nabla p^{n+1} + (fv, -fu, g\theta/\theta_0)^n - (F_u, F_v, 0)^n\right) \tag{117}$$

Imposing the condition $\nabla.\mathbf{u}^{n+1} = 0$ gives an equation exactly of the form (68), with the identification $p = s$. Thus the solution of (115) with a long timestep will tend exactly to the solution of (77), in the case where $(F_u, F_v) = -(c_D u, -c_D v)$. This achieves the aim of a solution of the general equations close to that of a reduced system if the time averaging scale is large, and also allows the subgrid model to be more completely accounted for than in (77). Thus a linearisation of the subgrid model only has to be included in the definition of the time tendency of the balanced state, not in that of the balanced state itself.

The identification of the long timestep solution with the specific semi-geostrophic reduced system (77) is a result of the explicit treatment of the Du/Dt term by the semi-Lagrangian algorithm. If the trajectory computation is made implicit, other reduced systems can be obtained as the long timestep limit. If the implicit calculation is simplified to use only a fixed reference θ profile which is independent of x and y, the quasi-geostrophic equations are obtained in the long timestep limit.

A further development of this algorithm would use the ideas of section 5.4 to identify the velocity calculated at time level $n+1$ with a total transport velocity $\mathbf{V}$ including sub-grid effects. This would require, for instance, allowing an implicit correction to be made to the convective mass transport. This can be achieved by writing the convection scheme as a nonlinear first estimate based on the values at the beginning of the timestep, together with a linearised correction.

7.3 Potential vorticity conservation

Now consider the accurate treatment of potential vorticity advection by this algorithm. The potential vorticity $\det\frac{\partial(X,Y,Z)}{\partial(x,y,z)}$ as given by (46) is conserved by the dynamical reduced equations (56), though not when the subgrid model is included. (X, Y, Z) are given by (31). Exact conservation of the potential vorticity following the motion is not practical using a conventional grid-point method. Contour dynamics algorithms, which use the potential vorticity as a variable and advect contours of fixed potential vorticity can achieve exact conservation, e.g. Dritschel and Ambaum (1997). It is possible to ensure conservation in the trajectory calculation by ensuring that the control volume associated with the

departure points corresponding to each grid-box has the same problem as the gridbox. This is illustrated in two dimensions in Fig. 16. We require in that case

$$\text{Area}(x_{1d}, z_{1d}, x_{2d}, z_{2d}) = (x_2 - x_1)(z_2 - z_1) \tag{118}$$

Conservation can also be improved by using the variables (X, Y, Z) directly in the integration scheme. Noting that Z is simply proportional to θ, (112) becomes

$$\begin{aligned}
f^{-1}(-Y,X)^{n+1} = \\
f^{-1}(-Y,X)^n_d - \delta t\left((-fv,fu)^{n+1} - (-fv,fu)^n + \nabla_h p^{n+1} - (F_u,F_v)^n\right) \\
w^{n+1} = w^n_d + \delta t\left(g\theta^{n+1}/\theta_0 - \frac{\partial p^{n+1}}{\partial z}\right) \\
\nabla.\mathbf{v}^{\mathbf{n+1}} = 0 \qquad (119) \\
\theta^{n+1} = \theta^n_d + \delta t\left(F^{n+1}_h + H - LP + S_h\right) \\
r^{n+1} = r^n_d + \delta t\left(F^{n+1}_r + P + S_r\right) \\
\mathbf{v}^{\mathbf{n+1}}.\mathbf{n} = 0 \text{ on } \partial\Gamma
\end{aligned}$$

Note that the implicit treatment of the Coriolis term is preserved. A similar procedure is analysed by Bates et al. (1995), where the implicit treatment of the Coriolis term is shown to be essential. Conservation is, however, lost in the interpolation to the departure points. In practice, this interpolation has to be done to higher order accuracy than the interpolation used in finding the departure point positions, so may be a less serious source of error. Other approaches to satisfying Lagrangian conservation laws in grid-point and spectral numerical methods are reviewed in the chapter by McLachlan later in this volume.

7.4 Grid design

The first design requirement stated in subsection 7.1 is to respect the fact that solutions of the full equations stay close to those of appropriate reduced systems. It is thus important that other motions, which are not part of the reduced solution, are not generated purely computationally, and that the numerical method can represent the solution of the reduced system accurately. If finite difference methods are used to implement the numerical procedure outlined above, it is important to ensure that the finite difference grid is suitable for representing the geostrophic and hydrostatic balance that defines the reduced system. Cullen (1989) describes a set of experiments to establish the best finite difference strategy for solving the semi-geostrophic system (56).

It is also important to ensure that implicit equations such as (115) can be solved easily in finite difference form. One consideration is the removal of zero eigenvalues from the discretised version of the Laplacian operator in (115). Another is to remove zero eigenvalues from the discretized version of $\mathbf{Q}$. It is always necessary to derive the discrete form of (115) from discrete approximations to (112). This can result in large stencils for approximating terms in (115), which lead to zero eigenvalues corresponding to 'checkerboard' modes, see Haltiner and Williams (1980, p.149, eq. 5-113).

The optimum grid for solving (113) is staggered so that u is a half grid length from p in the x direction, v in the y direction and w in the z direction (Fig. 17(a)). The horizontal

part of this arrangement is often referred to as the Arakawa C grid. This leads to the minimum possible stencil for representing $\nabla^2 p$. The optimum grid for dealing with the **Q** term in (115) can be deduced from the optimum grid for solving (68). This requires the components of ∇s to be a half gridlength from s in the appropriate direction. When applied to (115), the requirement is that v is a half grid length from p in the x direction, u in the y direction and θ in the z direction, illustrated in Fig. 17(b). The requirements for positions of u and v are thus contradictory. The staggering of θ from p in the vertical is known as the 'Charney-Phillips' grid. The advantages of this were demonstrated by Arakawa and Moorthi (1987).

In most atmospheric models using this type of integration scheme, the grid is chosen to optimise the representation of $\nabla^2 p$. However, in ocean models, where the resolution is often coarser compared with the scales of the motions being studied, the advantages of the two representations of u and v are more equal, and a compromise arrangement, the Arakawa 'B' grid is used (see Fig. 17(c) and Bryan (1989)). There are two other possible resolutions of the difficulty. The equivalence of u and v to ∇s only applies to the geostrophic part of the wind, which has no horizontal divergence. Instead of using u and v as variables, the horizontal divergence D and vertical component of vorticity ζ can be used. An ideal arrangement then positions both of these at the same point as p, Fig. 17(d). This is discussed by Bates et al (1995). The other possibility is based on the formulation of section 5.4 where there is an averaged velocity $\overline{\mathbf{u}}$ which is different from the total effective transport velocity **U**. The solution procedure would be ideal if these are held at the positions shown in Fig. 17(e). This arrangement has not yet been used in practice.

7.5 Example of implementation

We illustrate some of the above considerations by showing a comparison of potential vorticity forecasts from two versions of the UK Meteorological Office atmospheric model. The important features of the integration schemes are listed below.

Scheme A (Cullen and Davies (1991))

i) Variables held on Arakawa B grid in horizontal, θ held at same level as p.

ii) Explicit time integration.

iii) Advection using standard two step finite difference scheme with fourth order accuracy in space.

iv) Subgrid model in all equations includes horizontal eddy mixing term.

Scheme B (Cullen et al. (1997))

i) Variables held on Arakawa C grid in horizontal, Charney-Phillips grid used in vertical.

ii) Semi-implicit time integration as discussed above.

iii) Semi-Lagrangian advection, with monotonic interpolation for thermodynamic variables, but not for winds.

iv) No horizontal mixing in sub-grid model.

Scheme B includes several, but not all, of the desirable features discussed above. Comparative results after 10 days of integration are shown in Fig. 18. A latitude-longitude grid with 96x72 points was used, with 19 levels in the vertical. The output is for the level $\theta = 315K$. The potential vorticity distribution for scheme B shows much greater spatial coherence, even though the spatial smoothing has been removed from the sub-grid model. This is consistent with the discussion in section 6.4, and suggests that the design principles discussed in this section are important.

8 Application to data assimilation and predictability studies

8.1 Basic data assimilation techniques

Data assimilation is the process of finding a model state which best fits incomplete or imperfect data. The basic methods used in modern assimilation systems are reviewed by Daley (1997). We follow his definitions and notation in this section. The standard theory is developed for assimilation of data into a defined discrete model. The assumptions made about model behaviour and error structure are typically quite crude. In this section we show how some of the dynamical knowledge set out in earlier sections can be brought to bear on the problem.

8.2 Projection of observed information onto balanced states

Atmospheric observations are usually designed to sample the large scale low frequency motions assiciated with weather patterns. In section 4.5 of Daley (1997), it is pointed out that it is necessary to ensure that the data is projected onto these motions, and is not used to initialise small scale high frequency motions which can also be described by the equations. In practice, it is often necessary to use explicit procedures to damp unwanted high frequency motions, and force the data to be fitted by the low frequency motions, as illustrated in the chapter by Lynch in this volume.

The characterisation of the solution of a set of reduced equations as a sequence of minimum energy states described in sections 4.1 and 5.2 can be used to give a method of explicitly nudging a general state of a model solving equations (19) or (71) towards a state of geostrophic or geotriptic and hydrostatic balance. Given a general state $(\mathbf{u}, \theta, p)$, not satisfying (72), carry out a virtual displacement Ξ of the fluid defined by

$$\begin{aligned} \nabla^2 s &= \nabla.(fv - c_D u, -fu - c_D v, -g\theta/\theta_0) \\ \frac{\partial s}{\partial n} &= (fv - c_D u, -fu - c_D v, -g\theta/\theta_0).\mathbf{n} \text{ on } \partial\Gamma \\ \Xi \equiv (\xi, \eta, \zeta) &= \alpha((fv - c_D u, -fu - c_D v, -g\theta/\theta_0) - \nabla s) \end{aligned} \tag{120}$$

This ensures that $\nabla.\Xi = 0$ and $\Xi.\mathbf{n} = 0$ on $\partial\Gamma$. Use this displacement to change $(\mathbf{u}, \theta)$ on particles according to (73). Then the change to the balanced energy (44) can be shown to be

$$\delta e = \int_\Gamma -(f^2 + c_D^2)u^2 - (f^2 + c_D^2)v^2 - (g\theta/\theta_0)^2 - (\nabla s)^2 d\tau \qquad (121)$$

which is negative definite. Thus, with suitable choice of α, these displacements can be applied iteratively to give convergence to an energy minimising state satisfying (72). In practice, the solution for s from (120) will be very close to the hydrostatic pressure given by $\int g\theta/\theta_0 dz$. Then the components ξ and η of the displacement can be calculated explicitly from (120), and ζ from the condition $\nabla.\Xi = 0$. This procedure can be generalised to the free surface case. In the shallow water case, s is replaced by the fluid depth h, and the first two equations in (120) are not needed. The procedure can also formally be applied using the full sub-grid terms (F_u, F_v) in (120), though convergence cannot then be proved. It can be carried out globally, convergence is assured by the theorem of McCann quoted in section 3.6.

It can be shown that this iteration will converge fastest on the scales characterised by high frequency gravity waves. It will thus do a similar job to Lynch's digital filter, while ensuring that the potential vorticity is preserved on particles.

In order to try and project the observed information more directly onto the balanced part of the flow, it is convenient to change variables so that the primitive variables (u, v, θ) are replaced by three new variables, one of which contains all the balanced information, effectively the potential vorticity. A standard linear normal mode analysis of the hydrostatic version of (19) can be used to derive such a projection, as in Bartello (1995). θ has to be linearised about a basic state $\theta = \theta_0(z)$. The velocities are lineraised about asttae of rest. Writing the basic state Brunt-Vaisala frequency (9) as N^2, the first of the three new variables is the potential vorticity q, where

$$q = -f\frac{g}{\theta_0}\frac{\partial\theta}{\partial z} + N^2\left(\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}\right)$$

$$N^2\nabla_h^2 p + f^2\frac{\partial^2 p}{\partial z^2} = fq \qquad (122)$$

$$(u_1, v_1, \theta_1) = \left(-f^{-1}\frac{\partial p}{\partial y}, f^{-1}\frac{\partial p}{\partial x}, -\frac{\theta_0}{g}\frac{\partial p}{\partial z}\right)$$

The second is the ageostrophic part of the horizontally non-divergent wind and the pressure, such that the potential vorticity is zero. This is defined in terms of a stream-function Ψ satisfying

$$f^2\frac{\partial^2}{\partial z^2}\nabla_h^2\Psi + N^2\nabla_h^4\Psi = f\frac{\partial}{\partial z}\left(\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}\right) + \frac{g}{\theta_0}\nabla_h^2\theta \qquad (123)$$

$$(u_2, v_2, \theta_2) = \left(-f\frac{\partial^2\Psi}{\partial y\partial z}, f\frac{\partial^2\Psi}{\partial x\partial z}, \nabla_h^2\Psi\frac{\partial\theta_0}{\partial z}\right)$$

Substitution in the first equation of (122) confirms that this mode has $q = 0$. The third variable is the horizontal divergence, so that there is a velocity potential χ with

$$\nabla^2\chi = \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \qquad (124)$$

$$(u_3, v_3, \theta_3) = \left(\frac{\partial\chi}{\partial x}, \frac{\partial\chi}{\partial y}, 0\right)$$

This change of variables requires constant f and $\frac{\partial \theta_0}{\partial z}$. The second two components are not normal modes. The normal modes can only conveniently be written in Fourier space, as in Bartello (1995). The form written here is sometimes called the Craya-Herring cyclic basis. The projection can be generalised to spherical geometry with $\frac{\partial \theta_0}{\partial z}$ a general function of z by performing a spherical normal mode analysis, which cannot be written as a simple analytic procedure in real space. It is also possible to use (122) directly with f and $\frac{\partial \theta_0}{\partial z}$ allowed to vary. This gives so-called 'implicit normal mode' analysis, Temperton (1989).

We can alternatively use the energy minimisation procedure above to decompose a general state. This can be used in spherical geometry or in limited domains, and does not require any linearisations or restrictions on $\frac{\partial \theta}{\partial z}$. Given general (u, v, θ, p), use (120) to find a balanced state (u_b, v_b, θ_b, s), where s is the geostrophic and hydrostatic pressure, with the same 'balanced' potential vorticity, (46), as the original data. The original data can be reconstructed from (u_b, v_b, θ_b, s), given as functions of $\mathbf{x}$, and the integrated displacements Ξ, also given as functions of $\mathbf{x}$. The latter can be split into two components

$$\begin{aligned} \Xi &= \Xi_1 + \Xi_2 \\ \nabla_h \Xi_1 = 0, \zeta_1 &= 0 \\ \frac{\partial \eta_2}{\partial x} - \frac{\partial \xi_2}{\partial y} &= 0 \end{aligned} \tag{125}$$

Ξ_1 generates the horizontally divergent winds as in (124), and Ξ_2 generates the ageostrophic vorticity and temperature changes as in (123).

8.3 Initialisation of ageostrophic winds

In sections 4.4 and 5.3, we set out equations for the evolution of balanced states in the form (64). These include an equation for the total velocity $\mathbf{u}$, which does not satisfy the geostrophic relation, and acts as the response to dynamical and physical forcing. In section 5.4, it is shown that this total velocity can be interpreted as including the effect of subgrid transport. It includes, for example, the vertical motion balancing radiative heating or cooling. In order to analyse or initialise numerical forecasts, it is necessary to calculate this total velocity in a way which is consistent with the evolution of the balanced state.

The normal mode decomposition of the previous subsection projects all information onto the geostrophic and hydrostatic state. The total velocity consistent with the balanced state will also include components from the other normal modes. These components can be interpreted as being 'slaved' to the geostrophic state, see Warn et al. (1994) and the chapter by Warn later in this volume. They can be calculated by nonlinear normal mode initialisation, see Daley (1997), or other related methods. Nonlinear normal mode initialisation, including the physical forcing terms, is widely used in operational forecast models. The link between this procedure and solving an equation of the form (64) is made by Leith (1980). He shows that adiabatic nonlinear normal mode initialisation is equivalent to solving the quasi-geostrophic form of (64), with a fixed reference profile replacing the actual $\partial\theta/\partial z$. Use of (64) with (78) and (79) would then correspond to a variable coefficient generalisation of normal mode initilaisation, with physical forcing included.

It should be noted that a balance assumption is likely to be most accurate where the ageostrophic circulation required is small, and vice-versa. If forcing is applied to a state

with very small potential vorticity, (64) shows that the ageostrophic circulation required will be large, and possibly quite unrealistic. The real flow would simply be unbalanced under such conditions. Thus the obvious natural procedure of finding the balanced state that best fits the observations, and then calculating the associated total velocity from (64), is likely to be unsafe. The integration scheme (115) for the full equations avoids this problem by solving for $(\mathbf{I} + \mathbf{Q}\delta t^2)\mathbf{u}$, so that balance it not enforced when $\mathbf{Q}$ has a small eigenvalue. (115) should also thus be suitable as an intialisation scheme. This is consistent with the incremental nudging method of Lorenc et al. (1991), which effectively trusts the forecast model to initialise the total velocity field from information about the balanced state.

For the same reason, it is unwise to use the balanced ageostrophic motion as an error measure or in a cost function. The balanced pressure tendency is a safe variable to use instead, because the solution procedures discussed in section 5.3 will ensure that this evolves smoothly in time even when, as in deep convection, the balance requirement forces discontinuous mass transfers .

Many observations, such as satellite cloud observations, would be most useful as measures of the vertical motion. Using this data requires deducing the balanced state consistent with a given vertical motion. This is much harder, and leads to the use of the four-dimensional variational analysis method discussed below. The problem can be addressed by perturbing (64).

$$\mathbf{Q}\begin{pmatrix} u' \\ v' \\ w' \end{pmatrix} + \frac{\partial}{\partial t}\nabla s' = \mathbf{H}' \tag{126}$$

(127)

Taking the curl of (126) gives

$$\nabla \times \mathbf{Q}\begin{pmatrix} u' \\ v' \\ w' \end{pmatrix} = \nabla \times \mathbf{H}' \tag{128}$$

(129)

If $\mathbf{u}'$ is given, (128) allows $\mathbf{H}'$ to be calculated. If $\mathbf{H}$ is given by (79) a correction to the geostrophic flow can be calculated if the radiative forcing is assumed to be correct.

8.4 The use of Lagrangian methods to optimise the fit between two meteorological fields

One of the difficult issues in data assimilation is that the differences between two meteorological fields may not be well expressed in terms of a standard Eulerian norm such as the L_2 norm. This is because, at least away from equatorial regions, the evolution of the flow is dominated by advection of potential vorticity (section 4.5), and thus typical errors are likely to be displacements of the potential vorticity and hence the other fields derived from them. This is illustrated in Figs. 19 and 20. The root mean square position errors of

both extra-tropical and tropical cyclones grow approximately linearly with forecast time. The L_2 error of upper level height (equivalent to pressure) fields, which are dominated by large-scale and usually well forecast patterns also grows fairly linearly with time. However, the L_2 error of upper level winds, which are dominated by smaller scales features, increases most rapidly early in the forecast period and then saturates. The L_2 error of surface pressure, whose patterns are typically of smaller scale than upper air patterns, behaves in an intermediate fashion. Hoffman et al (1995) decomposed forecast errors of 500hpa height into displacement errors, amplitude errors, and a residual. They found that nearly 80 percent of the error in a particular 500hpa forecast could be explained as displacement error. The procedures of section 3.1 and 3.2 using rearrangements should also be applicable to this form of analysis.

Daley (1997) discusses the choice of norm to use in the analysis procedure in terms of expected distributions of errors. His equation (15) gives the total cost function to be minimised in the analysis as

$$J = 0.5\{[\mathbf{y}^o - H(\mathbf{x}^a)]^T \mathbf{R}^{-1} [\mathbf{y}^o - H(\mathbf{x}^a)] + [\mathbf{x}^f - \mathbf{x}^a]^T [\mathbf{P}^b]^{-1} [\mathbf{x}^f - \mathbf{x}^a]\} \qquad (130)$$

In (130), superscripts a, f, and o refer to the analysis, the first-guess forecast, and the observations respectively. $\mathbf{x}$ represents a model state, and $\mathbf{y}$ a set of observed values. The operator H calculates pseudo-observed values from forecast fields. The error measure to be used for the observations has to be related to the expected error sructure of the observations, encompassed by the error covariance matrix $\mathbf{R}$. This is discussed by Daley. The error measure to be used for the forecast minus analysis term has to be related to the expected forecast error structure, and so it is in this term that alternative norms such as measures of displacements could be used. The formulation (130) allows for the use of different norms in measuring observation error and forecast error, and produces a maximum likelihood estimate.

Hoffman et al. (1995) set out a procedure for minimising (130) where displacements are used in calculating the cost function. They also suggest using a displacement as a control variable in the minimisation, thus setting $\mathbf{x}_a$ equal to a displacement applied to $\mathbf{x}_f$ plus a residual. Given this approach, it is natural to seek a balanced fit to the data by using an additional displacement satisfying (73) to reduce the energy while preserving the potential vorticity, thus moving towards a geostrophic and hydrostatic state.

8.5 Sensitivity of model evolution to small perturbations - applications to data assimilation and ensemble generation

A working assumption for optimising the analysis procedure for numerical weather prediction is that the best analysis is that which leads to the best forecast. therefore it is particularly important to analyse accurately the features which could lead to the fastest error growth in the subsequent forecast, see Courtier (1997). If we use the idea set out in section 4.5 that the atmosphere stays close to an evolution given by advection of potential vorticity, then information about the sensitive regions can be deduced from a linear perturbation of (69):

$$\frac{\partial q'}{\partial t} + \mathbf{U}.\nabla q' + \mathbf{U}'.\nabla q \quad = 0 \qquad (131)$$

$$\mathcal{H}(\mathbf{U}', \theta') = q'$$

Maximum sensitivity is likely to occur when $\mathbf{U}$ is normal to ∇q and both are large. Then a fixed perturbation to either will have the largest effect on $\mathbf{U}.\nabla q$. According to the principles reviewed in section 6.2, these will be steady states corresponding to stationary but non-extremal points of the energy under potential vorticity rearrangements. In a barotropic flow, these are characterised locally by points where q is non-monotonic as a function of streamfunction. The degree of sensitivity can be inferred from the energy difference between the actual profile and that that would be obtained from a local monotone rearrangement of q. Baroclinic steady flows will additionally correspond to non-extremal points of the energy if the vertical correlation of the potential vorticity gradient on isentropic surfaces can be increased. Buizza and Palmer (1995) illustrate a correlation between sensitive regions and the local value of du/dz, which is a crude measure of the potential for baroclinic instability.

Another contributing factor will be a small eigenvalue of $\mathcal{H}^{-1}$, which will give a large $\mathbf{U}'$ for a given q'. If the potential vorticity inversion is given by (64) and (68), then we have

$$\begin{aligned}
\mathbf{Q}\begin{pmatrix} u' \\ v' \\ w' \end{pmatrix} + \frac{\partial}{\partial t}\nabla s' &= \mathbf{H}' \\
\nabla.\mathbf{u}' &= 0 \\
(fv_b', -fu_b', g\theta'/\theta_0) &= \nabla s' \\
\nabla.\mathbf{Q}^{-1}\frac{\partial}{\partial t}\nabla s' &= \nabla.\mathbf{Q}^{-1}\mathbf{H}' \\
(\mathbf{Q}^{-1}\frac{\partial}{\partial t}\nabla s')_n &= (\mathbf{Q}^{-1}\mathbf{H}')_n \text{ on } \partial\Gamma
\end{aligned} \tag{132}$$

It can be seen from the last of these equations that small eigenvalues of the potential vorticity matrix $\mathbf{Q}$ will give the greatest sensitivity. Note also that the perturbation has to be put in in terms of the forcing matrix $\mathbf{H}$. This shows that the natural way the effects of forcing are spread in space depend on the structure of the $\mathbf{Q}$ matrix. This has been applied in designing data assimilation systems by stating that the region of influence of an observation should be isotropic in geostrophic and isentropic coordinates, see Desroziers and Lafore (1993).

There are two methods commonly used in operational centres of calculating sensitivity directly from a numerical model solving the evolution equations. The first is the 'singular vector' method. This is expensive to use, and is thus normally used with a reduced set of equations. A set of equations for the evolution of a perturbation to the original forecast is derived by linearising the governing equations. If these take the form

$$\frac{\partial \mathbf{u}}{\partial t} = N\mathbf{u} \tag{133}$$

the 'tangent linear' equations for a perturbation to the solution takes the form

$$\frac{\partial \mathbf{u}'}{\partial t} = L(t)\mathbf{u}' \tag{134}$$

where the operator $L(t)$ depends on the control forecast. This system is written in discrete form, when it becomes a matrix equation with time-dependent coefficients. If this is integrated over a fixed number of timesteps, it generates a single matrix evolution operator

$$\mathbf{u}'(t) = E(0,t)\mathbf{u}'(0) \tag{135}$$

The eigenvalues of $E(0,t)$ are calculated, and that giving the largest growth of $\mathbf{u}'$ in a suitable norm, usually energy, is chosen. The corresponding eigenfunction gives the structure of the most sensitive perturbation to the initial state. The procedure is fully described by Buizza and Palmer (1995).

The second method is the 'error-breeding method' of Toth and Kalnay (1993). This consists of finding the structure of the fastest growing perturbation to a given initial state iteratively by integrating a perturbed forecast for a few hours, calculating the difference from the control forecast, and then repeating the exercise with an initial perturbation equal to the rescaled final perturbation from the previous iteration. This procedure converges to the mode with the largest Liapunov exponent. It would be interesting to compare the results of both these methods with the dynamically based methods discussed at the beginning of this subsection.

It remains to be explored whether the areas of maximum sensitivity identified by the two latter methods can be characterised in terms of the properties of the potential vorticity distribution set out above. If so, local potential vorticity diagnostics could be used to change the assumed error in the first guess forecast, and/or to change the overall weighting of the cost function in (130) so that the analysis procedure is focussed towards doing a good job in critical areas.

8.6 Fitting observations by a model trajectory over a time interval

As noted above, 4D variational data assimilation is the most promising avenue for using observations relating to the ageostrophic circulation. In addition, as discussed by Eyre (1997), it allows use of observations with a high time frequency, such as data from aircraft, and gives a framework for use of remotely sensed data which is only indirectly related to the variables predicted directly by the model. Daley (1997) shows that the theoretical basis for 4DVAR comes from the need to predict the evolution of model error, and is a compromise from the more complete method based on the Kalman filter forced by lack of computer power. However, even the Kalman filter method assumes that model errors are uncorrelated in time, which is extremely unlikely. The hope is that more dynamical knowledge can be used to replace these statistical assumptions.

The standard 4DVAR strategy is to minimise the cost function (130) over a fixed time interval. The output is a model trajectory which satisfies the (discrete) equations and minimises (130). Since the model is deterministic, the control variables have to be the initial values of the model variables. In practice, this minisation has to be done iteratively using the linearisation (134) of the model equations, and their adjoint (see Daley (1997)). A number of additional possibilities are suggested by the preceding sections:

a) Use of a model integration scheme based on the principles of section 7 to ensure that the model trajectory stays close to a balanced state. In particular, use of (115) for initialisation, and use in the tangent linear model (134) of the same linearised corrections to

the subgrid model as are needed for the implicit time integration.
b) Use of potential vorticity, divergence and geostrophic departure as control variables, so that more use can be made of dynamical information in estimating the error structure of the potential vorticity. Overall estimates of the statistics of the unbalanced motion can be made by the methods of section 4.2.
c) Use of displacement error in calculating the difference between two forecast fields, or between a forecast field and an analysed field derived from a set of observations. Use of displacements in the iteration to minimise the cost function, incorporating the energy minimisation method in the iterative procedure to aid convergence to a near balanced state.
d) Use of the potential vorticity matrix in estimating how to spread information, or the likely structure of the error matrix.
e) Use of dynamical estimates of sensitivity in estimating model error, or in the overall weighting of the cost function to be minimised.

In addition, a basic and well recognised limitation of the current 4DVAR approach is the need to treat the model as perfect, and hence use the initial values as control variables. An alternative would be to use the model as a weak constraint. The simplest way of doing this would be to have a complete set of model variables as control variables at a number of times during the assimilatipon period. This is very expensive and not practical. Courtier (1997) and Daley (1997) review a number of ways of achieving this in an affordable way. Since the operational forecasting requirement is to get the most accurate estimate of the atmospheric state at the end of the assimilation period, it is natural to consider using the final values rather than the initial values as control variables. In this way the lack of knowledge of model imperfections should be less serious for the final product. The difficulty is that the full equations of motion with sub-grid model used for the forecast cannot be sensibly integrated backwards in time.

There are two approaches which could allow construction of the analysis procedure back from the endpoints. If only small changes are expected, so that the 'tangent linear' model (134) correctly describes the evolution of the error, then it is possible to construct a time-reversed tangent linear model and use that in essentially the same way as the normal algorithm. Alternatively, we can use the close approximation by a reduced set of equations and the fact that (64) can be time reversed in the absence of the sub-grid model. (Note that tim ereversal for a long period requires the result of section 6.4 that there si no enstrophy cascade in certain types of reduced equations). The theory governing these equations set out in the earlier sections 3.5, 4.3 and 5.3 shows that this full time reversibility applies up to points where discontinuous convective mass transfers occur. These will be places where the same final state can result from several initial states, and thus the influence of observations will be lost.

9 References

ABARBANEL, H.D.I. and HOLM, D.D. (1987) Nonlinear stability analysis of inviscid flows in three dimensions: Incompressible fluids and barotropic fluids. *Phys. Fluids*, **30**, 3369-3382.

ALLEN,J.S., BARTH,J.A. and NEWBERGER,P. (1990) On intermediate models for barotropic continental shelf and slope flow fields. Part I: Formulation and comparison of

exact solutions.. *J. Phys. Oceanog.*,**20**,1017-1043.

ARAKAWA,A. and MOORTHI,S. (1987) Baroclinic instability in vertically discrete systems. *J. Atmos. Sci.* **45**, 1688-1707.

BABIN,A., MAHALOV,A. and NICOLAENKO, B. (1996) Global splitting, integrability and regularity of 3D Euler and Navier-Stokes equations for uniformly rotating fluids. *Eur. J. Mech.,B/Fluids*,**15**, 291-300.

BABIN,A., MAHALOV,A. and NICOLAENKO, B. (1997) Regularity and integrability of rotating shallow water equations. *C.R.Acad.Sci. Paris*,**324**, 593-598.

BARTELLO,P. (1995) Geostrophic adjustment and inverse cascades in rotating stratified turbulence. *J. Atmos. Sci.*,**52**, 4410-4428.

BARTELLO,P. and THOMAS,S.J. (1996) The cost-effectiveness of semi-Lagrangian advection. *Mon. Weather Rev.*,**124**, 2883-2897.

BATES,J.R., LI,Y., BRANDT,A., MCCORMICK,S.F. and RUGE,J. A global shallow water numerical model based on the semi-Lagrangian advection of potential vorticity. *Quart. J. Roy. Meteor. Soc.*, **121**, 1981-2006.

BRENIER,Y. (1990) The least action principle and the related concept of generalised flows for incompressible inviscid fluids. *J. Amer. Math. Soc.*, **2**, 225-255.

BRENIER,Y. (1991) Polar factorization and monotone rearrangement of vector-valued functions. *Commun. Pure Appl. Math.*, **44**, 375-417.

BRYAN,K. (1989) The design of numerical models of the ocean circulation in Oceanic circulation models: combining data and dynamics. Proceedings of the NATA Advanced Study Institute, Les Houches, France, February 15-26, 1988 ,465-500.

BUHLER,O. (1996) Waves and balanced mean flows in the atmosphere. Ph. D. thesis, Department of Appl. Math. and Theor. Phys., Univ. of Cambridge, U.K., 183pp.

BUIZZA,R. and PALMER,T.N. (1995) The singular-vector structure of the atmospheric global circulation. *J. Atmos. Sci.*,**52**, 1434-1456.

BURTON,G.R. and DOUGLAS,R.J. Rearrangements and polar factorisation of countably degenerate functions.

BURTON,G.R. and MCLEOD,J.B. (1991) Maximisation and minimisation on classes of rearrangements. *Proc. Roy. Soc. Edin.* (A),**119**,287-300.

BURTON,G.R. and NYCANDER,J. (1996) Stationary vortices in three-dimensional quasi-geostropic shear flow. submitted

CAFFARRELLI,L.A. (1996) Boundary regularity of maps with convex potentials. *Annals of Math.*,**143**, 1-43.

CHARNEY, J.G., FJORTOFT, R. and VON NEUMANN,J. (1950) Numerical integration of the barotropic vorticity equation. *Tellus*, **2**, 237-254.

COURTIER,P. (1997) Variational methods. *J. Met. Soc. Japan*, **75**, 211-218.

CULLEN,M.J.P. (1989) Implicit finite difference methods for modelling discontinuous atmospheric flows. *J. Comp. Phys.*, **81**, 319-348.

CULLEN,M.J.P. and DAVIES,T. (1991) A conservative split-explicit integration scheme with fourth order horizontal advection. *Quart. J. Roy. Meteor. Soc.*,**117**, 993-1002.

CULLEN,M.J.P., DAVIES,T., MAWSON,M.H., JAMES,J.A. and COULTER,S. (1997) An overview of numerical methods for the next generation UK NWP and climate model. *Atmos.-Ocean*, Andre Robert Memorial Symposium issue, to appear.

CULLEN,M.J.P., DOUGLAS,R.J. and ROULSTONE,I. (1997) Generalised semi-geostrophic theory on a sphere. In preparation.

CULLEN,M.J.P. and MAWSON,M.H. (1992) An idealized simulation of the Indian monsoon using primitive-equation and quasi-equilibrium models. *Quart. J. Roy. Meteor. Soc.*,**118**,153-164.

CULLEN,M.J.P., NORBURY,J., PURSER,R.J. and SHUTTS,G.J. (1987) Modelling the quasi-equilibrium dynamics of the atmosphere. *Quart. J. Roy. Meteor. Soc.*,**113**,735-757.

CULLEN,M.J.P. and PURSER,R.J. (1984) An extended Lagrangian theory of semi-geostrophic frontogenesis. *J. Atmos. Sci.*,**41**,1477-1497.

CULLEN,M.J.P. and PURSER,R.J. (1989) Properties of the Lagrangian semi-geostrophic equations. *J. Atmos. Sci.*,**46**,2684-2697.

DALEY,R. (1997) Atmospheric data assimilation. *J. Met. Soc. Japan*, **75**, 319-329.

DESROZIERS,G. and LAFORE,J.-P. (1993) A coordinate transformation for objective frontal analysis. *Mon. Weather Rev.*, **121**, 1531-1553.

DIPERNA,R.J. and LIONS,P-L.L. (1989) Ordinary differential equations, transport theory and Sobolev spaces. *Invent. Math.*,**98**, 511-547.

DOUGLAS,R.J. (1994) Rearrangements of functions on unbounded domains. *Proc. Roy. Soc. Edin.*(A),**124**, 621-644.

DRITSCHEL,D.G. and AMBAUM,M.H.P. (1997) A contour-advective semi-Lagrangian numerical algorithm for simulating fine-scale conservative dynamical fields. *Quart. J. Roy. Meteor. Soc.*, **123**, 1097-1130.

EMANUEL,K.A. (1983) The Lagrangian parcel dynamics of moist symmetric instability. *J. Atmos. Sci.*, **40**, 2369-2376.

EMANUEL,K.A., NEELIN, J.D. and BRETHERTON, C.S. (1994) On large-scale circulations in convecting atmospheres. *Quart. J. Roy. Meteor. Soc.*,**120**, 1111-1144.

EMBID,P.F. and MAJDA,A.J. (1996) Averaging over fast waves for geophysical flows with arbitrary potential vorticity. *Comm. Partial Diff. Eqs.*, **21**, 619-658.

EYDELAND,A.,SPRUCK,J. and TURKINGTON,B. (1990) Multiconstrained variational problems of nonlinear eigenvalue type: new formulations and algorithms. *Math. Comp.* **55**, 509-535.

EYRE,J.R. (1997) Variational assimilation of remotely-sensed observations of the atmosphere. *J. Met. Soc. Japan*, **75**, 331-338.

FJORTOFT (1953) On the changes in the spectral distribution of kinetic energy for two-dimensional, non-divergent flow. *Tellus*, **5**, 225-230.

GAGE,K.S. and NASTROM,G.D. (1986) Theoretical interpretation of atmospheric wavenumber spectra of wind and temperature observed by commercial aircraft during GASP. *J.Atmos. Sci.*, **43**, 729-740.

GANGBO,W. and R.J. MCCANN,R.J. (1997) The geometry of optimal transportation.*Acta Math.*, **177:2**, 113-161.

GARRETT,C and MUNK,W. (1979) Internal waves in the ocean. *Ann. Rev. Fluid Mech.*,**11**, 339-369.

GENT,P.R. and McWilliams,J.C. (1996) Eliassen-Palm fluxes and the momentum equation in non-eddy-resolving ocean circulation models. *J. Phys. Oceanog.*,**26**,2539-2546.

GERARD,P. (1992) Resultats recents sur les fluides parfaits incompressibles bidimensionnels. *Seminaire Bourbaki*, **757**, 411-444.

GILL,A.E. *Atmosphere-Ocean Dynamics.* Academic Press. 662pp.

GJAJA,I. AND HOLM,D.D. (1996) Self-consistent hamiltonian dynamics of wave, mean-f;low intercation for a rotating startiufied incompressible fluid. *Physica D*, **98**, 343-378.

HALMOS,P.R. (1950) *Measure Theory.* Van Nostrand, New York

HALTINER, G.J. and WILLIAMS,R.T. (1980) *Numerical prediction and Dynamic Meteorology* 2nd ed. John Wiley and Sons. 477pp.

HOLTON, J.R. (1992) *An introduction to dynamic meterology.* Academic Press, 511pp.

HOFFMAN, R.N.,LIU, Z., LOUIS, J-F. and GRASSOTTI, C. (1995) Distortion representation of forecast errors.*Mon. Wea. Rev.*, **123**, 2758-2770.

HOSKINS,B.J. (1975). The geostrophic momentum approximation and the semi-geostrophic equations. *J. Atmos. Sci.*,**32**,233-242.

HOSKINS,B.J., MCINTYRE,M.E. and ROBERTSON,A.W. (1985) On the use and signifance of isentropic potential vorticity maps. *Quart. J. Roy. Meteor. Soc.*,**111**,887-946.

JOHNSON,W.B. (1966) The geotriptic wind. *Bull. Amer. meteor. Soc.*, **47**, 982.

KATO,T. and PONCE,G. (1986) Well posedness of the Euler and Navier-Stokes equations in Lebesgue spaces $L^p_s(\mathbb{R}^2)$. *Rev. Mat. Iberoamerica*,**2**, 73-88.

KNOX,J.A. (1997) Generalized nonlinear balance criteria and inertial stability. *J. Atmos. sci.*, **54**, 967-985.

KUSHNER,P.J. and SHEPHERD,T.G. (1995) Wave-activity conservation laws and stability theorems for semi-geostrophic dynamics. Part 2. Pseudo-energy based theory. *J.Fluid Mech.*, **290**,105-129.

LEITH,C.E. (1980) Nonlinear normal mode initialization and quasi-geostrophic theory. *J.Atmos.Sci.*,**37**, 958-968.

LEITH,C.E. (1983) Predictability in theory and practice. *Large-scale dynamical processes in the atmosphere*, ed. B.J.Hoskins and R.P.Pearce, Academic Press, pp. 365-383.

LORENC,A.C., BELL,R.S. and MACPHERSON,B. (1991) The Meteorological Office analysis correction data assimilation scheme. *Quart. J. Roy. Meteor. Soc.*, **117**, 59-90.

LORENZ,E.N. (1963) Deterministic nonperiodic flow. *J. Atmos. Sci.*, **20**, 130-141.

MARSDEN,J.E., RATIU,T.S. and Raugel,G. (1995) Equations d'Euler dans une coque spherique mince. *C.R.Acad. Sci. Paris*, **321**, 1201-1206.

NASTROM,G.D. and GAGE,K.S. (1985) A climatology of atmospheric wavenumber spectra observed by commercial aircraft. *J. Atmos. Sci.*, **42**, 950-960.

NEVEN,E.C. (1994) Baroclinic modons on a sphere. *J. Atmos. Sci.*,**51**, 1447-1464.

OLSSON,P.Q. and COTTON,W.R. (1997) Balanced and unbalanced circulations in a primitive equation simulation of a midlatitude MCC. Part II. Analysis of balance. *J. Atmos. Sci.*,**54**, 457-497.

OSTDIEK,V. and BLUMEN,W. (1997) A dynamic trio: inertial oscillation, deformation frontogenesis, and the Ekman-Taylor boundary layer. *J. Atmos. Sci.*, **54**, 1490-1502.

PIELKE, R.A. (1984) *Mesoscale Meteorology*, Academic Press, 612pp.

POGORELOV,A.V. (1973) *Extrinsic geometry of convex surfaces*, Israel Progamme for Scientific Translations, 669pp.

PURSER,R.J. (1996) Legendre-transformable semi-geostrophic theories. *J. Atmos. Sci.*,,submitted.

RICHARDSON, L.F. (1922). *Weather Prediction by Numerical Process.* Cambridge

University Press, reprinted by Dover, 1965. 236pp.

ROULSTONE,I. and MCINTYRE,M.E. (1996). Hamiltonian balanced models: constraints, slow manifolds and velocity splitting. *J. Fluid Mech.*. ,submitted.

ROULSTONE,I. and SEWELL,M.J. (1996). Potential vorticities in semi-geostrophic theory. *Quart. J. Roy. Meteorol. Soc.*,**122**,983-992.

SALMON,R. (1985) New equations for nearly geostrophic flow. *J.Fluid Mech.*,**153**,461-477.

SCHUBERT,W.H. (1985) Semi-geostrophic theory. *J. Atmos. Sci.*,**42**, 1770-1772.

SCHUBERT,W.H., CIESIELESKI,P.E., STEVENS,D.E. and KUO H-C (1991) Potential vorticity modelling of the ITCZ and the Hadley circulation. *J. Atmos. Sci.*,**48**,1493-1509.

SHUTTS,G.J. (1987) Balanced flow states resulting from penetrative, slantwise convection. *J. Atmos. Sci.*, **44**, 3363-3376.

SHUTTS,G.J. (1987) The semigeostrophic weir: a simple model of flow over mountain barriers. *J. Atmos. Sci.*, **44**, 2018-2030.

SHUTTS,G.J. (1995) An analytical model of the balanced flow created by localized convective mass transfer in a rotating fluid. *Dyn. Atmos. Oceans*, **22**,1-17.

SHUTTS,G.J. and CULLEN,M.J.P. (1987) Parcel stability and its relation to semi-geostrophic theory. *J. Atmos. Sci.*,**46**, 2684-2697.

SHUTTS,G.J., BOOTH,M. and NORBURY,J. (1988) A geometric model of balanced, axisymmetric flow with embedded penetrative convection. *J. Atmos. Sci.*,**45**, 2609-2621.

SMAGORINSKY, J. (1974) Global atmospheric modeling and the numerical simulation of climate., in 'Weather and Climate Modification', John Wiley and Sons, New York, 842pp, pp 633-686.

SNYDER,C., SKAMAROCK,W.C. and ROTUNNO,R. (1991) A comparison of primitive-equation and semi-geostrophic simulations of baroclinic waves. *J. Atmos. Sci.*,**48**, 2179-2194.

STANIFORTH,A. and COTE,J. (1991) Semi-Lagrangian integration schemes for atmospheric models - A review. *Mon. Weather Rev.*,**119**, 2206-2223.

TEMPERTON,C. (1989) Implicit normal mode initialisation for spectral models. *Mon. Weather Rev.*, **117**, 436-451.

THOMSON,W. (LORD KELVIN) (1910) Maximum and minimum energy in vortex motion. *Mathematical and Physical Papers*, vol. 4 172-183, Cambridge University Press.

THORPE,A.J. and Emanuel,K.A. (1985) Frontogenesis in the presence of small stability to slantwise convection. *J. Atmos. Sci.*,**42**, 1809-1824.

TOTH,Z. and KALNAY E. (1993) Ensemble forecasting at NMC: the generation of perturbations. *Bull Amer. Meteor. Soc*, **74**, 2317-2330.

VALLIS,G.K. (1992) Mechanisms and parametrizations of geostrophic adjustment and a variational approach to balanced flow. *J. Atmos. Sci.*,**49**, 1144-1160.

VALLIS,G.K. (1996) Mechanisms and parametrizations of geostrophic adjustment and a variational approach to balanced flow. *Quart. J. Roy. Meteor. Soc.*,**122**, 291-322.

VEITCH,G. and MAWSON,M.H. (1993) A comparison of inertial stability conditions in the planetary semi-geostrophic and quasi-equilibrium models. U.K. Met. Office Short Range Forecasting Tech. Report no. 60.

WARN,T., BOKHOVE,O., SHEPHERD,T.G. and VALLIS,G.K. (1994) Rossby number expansions, slaving, and balance dynamics. *Quart. J. Roy. Meteor. Soc.*,**121**, 723-739.

10 Figure captions

Figure 1. Summary components of wavenumber spectra of potential temperature and zonal and meridional velocity, composited from three groups of flight segments of different lengths. The 3 types of symbol show results from each group. The straight lines indicate slopes of -3 and -5/3. The meridional wind spectra are shifted one decade to the right and the potential temperature spectra are shifted two decades to the right (after Nastrom and Gage, 1985) .

Figure 2. Typical space and time scales of atmospheric phenomena (following Smagorinsky (1974)).

Figure 3(a) METEOSAT picture for 1500 UTC on 13 March 1995.
(b) High resolution picture for 1315 UTC on 13 March 1995.

Figure 4. Frequency histogram of inertial stability deduced from automated aircraft reports between 30N and 30S using dates between June 1992 and June 1993. The parameter β plotted is defined as $1 + f^{-1}\zeta$, where ζ is the relative vorticity. Inertial stability requires $\beta > -1$.

Figure 5. Definition of main asymptotic regimes in the atmosphere and ocean presented as plot against Froude number Fr and 'anisotropic' Rossby number $Ro_a = Ro/A$. QG refers to balanced motion, and AG to inertio-gravity waves.

Figure 6. Graphs of $f(x) = x, x \in [0,1]$ and

$$g(x) = \begin{cases} 1 - 2x & \text{if } x \in [0, 1/2], \\ 2x - 1 & \text{if } x \in [1/2, 1]. \end{cases} \tag{136}$$

illustrating the reason why f is a rearrangement of g.

Figure 7. a) Radar rainfall observations over the UK at 1800 UTC on 20 February 1997.
b) 18 hour numerical model forecast rainfall valid at 1800 UTC on 20 February 1997.

Figure 8. Construction of a convex polyhedral surface $S(x, z)$ from faces with given gradients (X_i, Z_i) and areas .

Figure 9. The rearrangements $f_3 = f_3(x)$ and $f_8 = f_8(x)$ defined by equation (42) of the function $f_0 = f_0(x)$ defined by equation (41).

Figure 10. Illustration of the comparison between the error of a prediction $u(t)$ using a reduced set of equations, and the projection P of the exact solution $U(t)$ onto a state consistent with the reduced system.

Figure 11. Forecast wind vectors at surface from idealised simulation of cross-equatorial flow using (a) equations (77), (b) equations (71).

Figure 12. Daily weather maps from 1-30 April 1997. Dates are ringed at lower left hand corner of each map. Maps are of sea level pressure with contour interval of 4hpa. Source *Weather Log*, Royal Meteorological Society.

Figure 13. (a) ρ distribution corresponding to a steady baroclinic shear flow. (b) The minimum energy rearrangement of this distribution. (c) The minimum energy configuration for a ρ distribution which cannot be made indepenedent of Z.

Figure 14. Three example distributions of the inverse potential vorticity for a semi-geostrophic shear flow.

Figure 15. Distributions in $\mathbf{x}$ and $\mathbf{X}$ space corresponding to steady solution of semi-geostrophic equations on a sphere.

Figure 16. Control volumes for ensuring conservation of potential vorticity using semi-Lagrangian advection.

Figure 17. Various arrangements of variables on the grid for solution of (115). (a) Natural staggering to give best stencil for $\nabla^2 p$. (b) Natural staggering for representing ∇s. (c) Arakawa B grid. (d) Natural arrangement using D and ζ as variables. (e) Natural arrangement using $\overline{\mathbf{u}}$ and $\mathbf{U}$ as variables.

Figure 18. 10 day forecasts of potential vorticity at 315K using the Meteorological Office Unified Model in a climate model configuration and integration scheme A (left) and the same with integration scheme B (right).

Figure 19. Growth of rms position error with forecast time for extratropical and tropical cyclones in the Northern hemipshere during 1995.

Figure 20. Growth of L_2 error of surface pressure, 500hpa height, and 200hpa wind over the Northern hemisphere with forecast time during 1995.

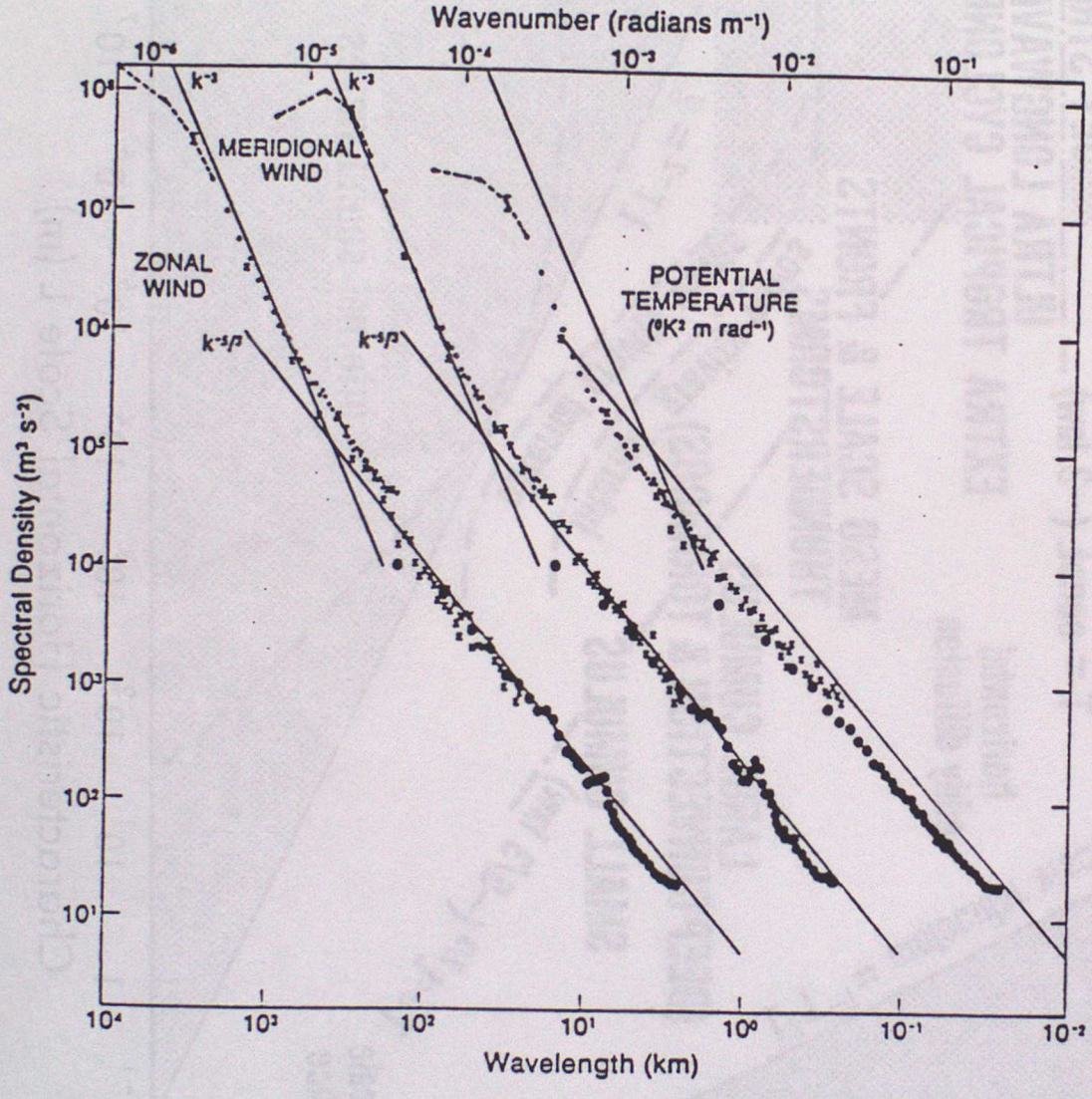
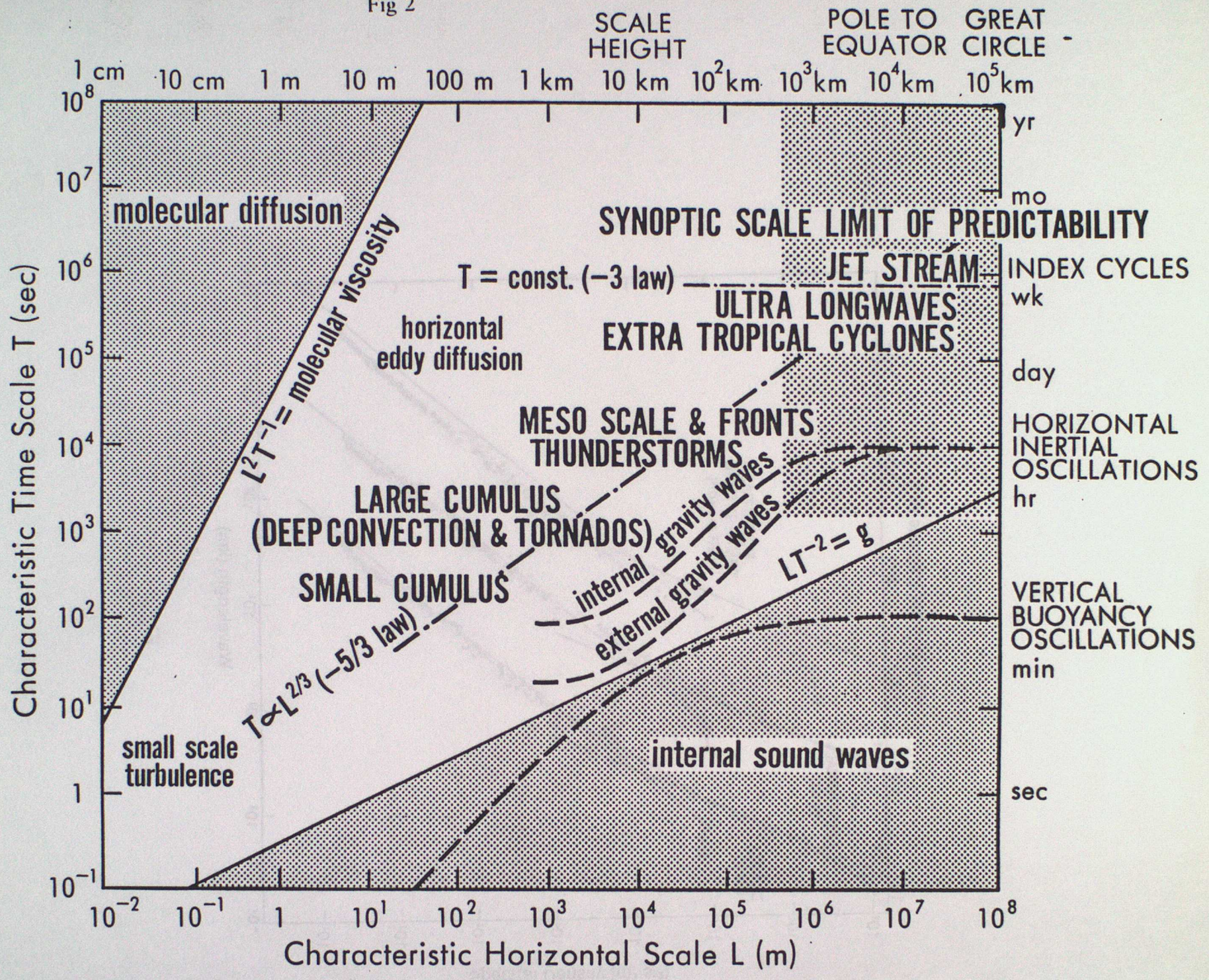

Wavenumber (radians m⁻¹)
10⁻⁶
10⁻⁵
10⁻⁴
10⁻³
10⁻²
10⁻¹
k⁻³
k⁻³
MERIDIONAL WIND
ZONAL WIND
POTENTIAL TEMPERATURE (°K² m rad⁻¹)
k⁻⁵/³
k⁻⁵/³
Spectral Density (m³ s⁻²)
10⁸
10⁷
10⁶
10⁵
10⁴
10³
10²
10¹
10⁴
10³
10²
10¹
10⁰
10⁻¹
10⁻²
Wavelength (km)

Fig 1

Fig 2


SCALE HEIGHT
POLE TO EQUATOR GREAT CIRCLE
1 cm
10 cm
1 m
10 m
100 m
1 km
10 km
10^2km
10^3km
10^4km
10^5km
10^8
10^7
10^6
10^5
10^4
10^3
10^2
10^1
1
10^{-1}
yr
mo
wk
day
hr
min
sec
molecular diffusion
L^2T^{-1} = molecular viscosity
horizontal eddy diffusion
SYNOPTIC SCALE LIMIT OF PREDICTABILITY
T = const. (−3 law)
JET STREAM
ULTRA LONGWAVES
EXTRA TROPICAL CYCLONES
INDEX CYCLES
MESO SCALE & FRONTS THUNDERSTORMS
HORIZONTAL INERTIAL OSCILLATIONS
LARGE CUMULUS (DEEP CONVECTION & TORNADOS)
SMALL CUMULUS
$T \propto L^{2/3}$ (−5/3 law)
internal gravity waves
external gravity waves
$LT^{-2} = g$
VERTICAL BUOYANCY OSCILLATIONS
small scale turbulence
internal sound waves
Characteristic Time Scale T (sec)
10^{-2}
10^{-1}
1
10^1
10^2
10^3
10^4
10^5
10^6
10^7
10^8
Characteristic Horizontal Scale L (m)

METEOSAT VIS 13 MAR 1995 15:00 UTC

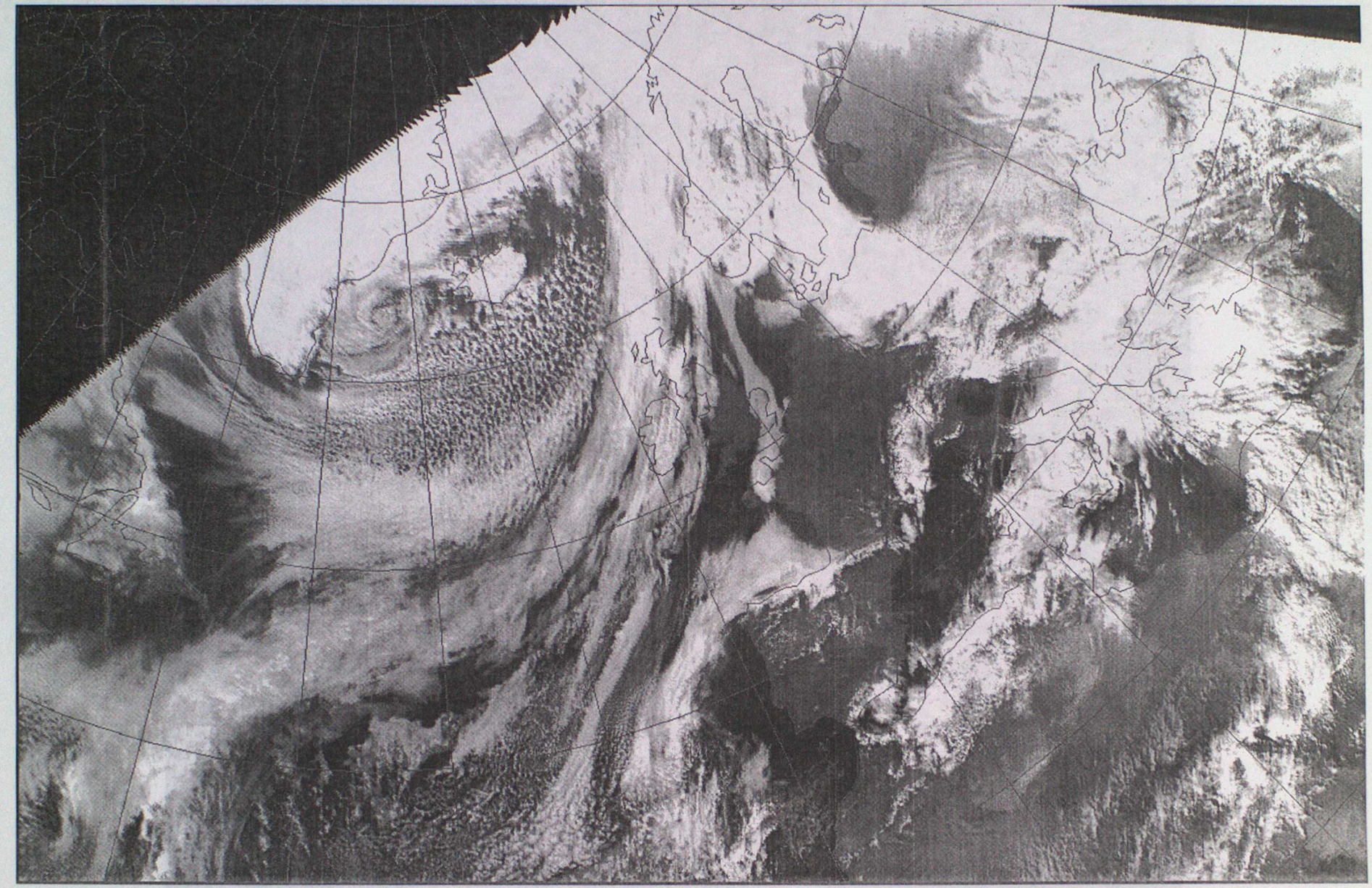


Fig 3a

AVHRR VIS 13 MAR 1995 13:15 UTC

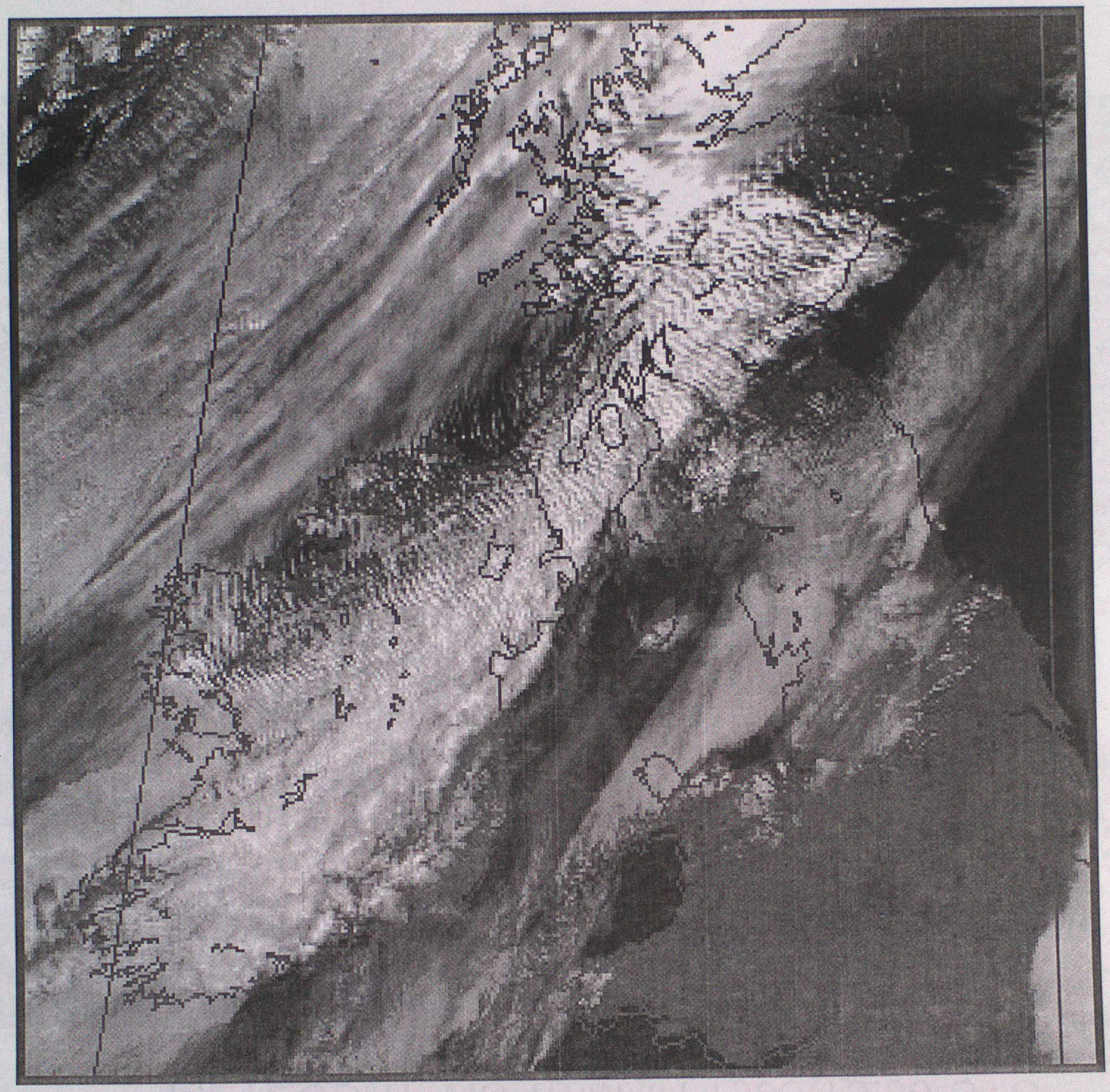


Fig 3b

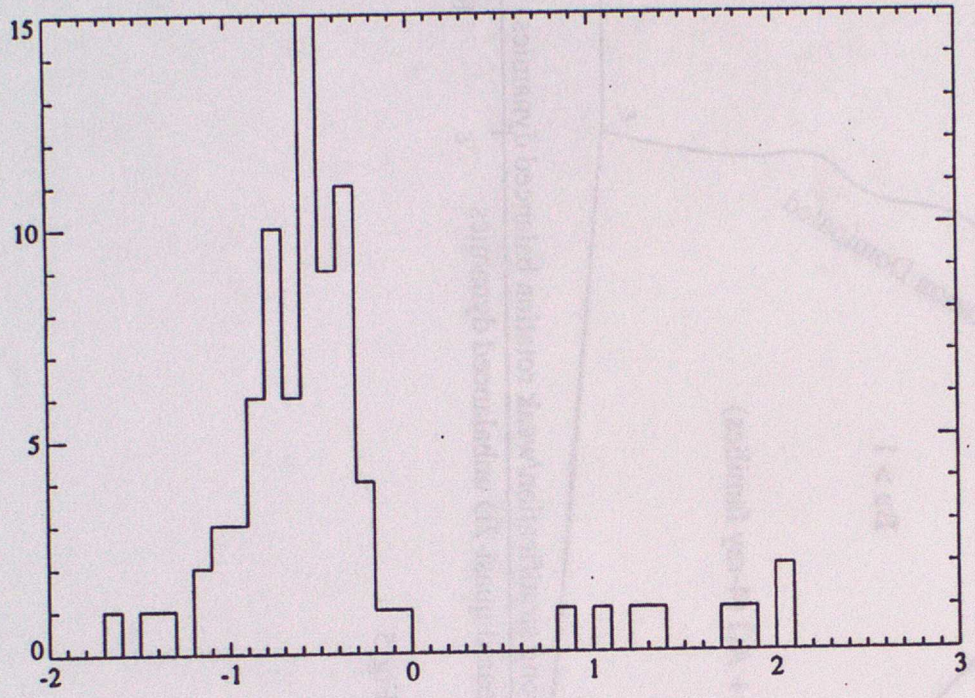


Fig 4

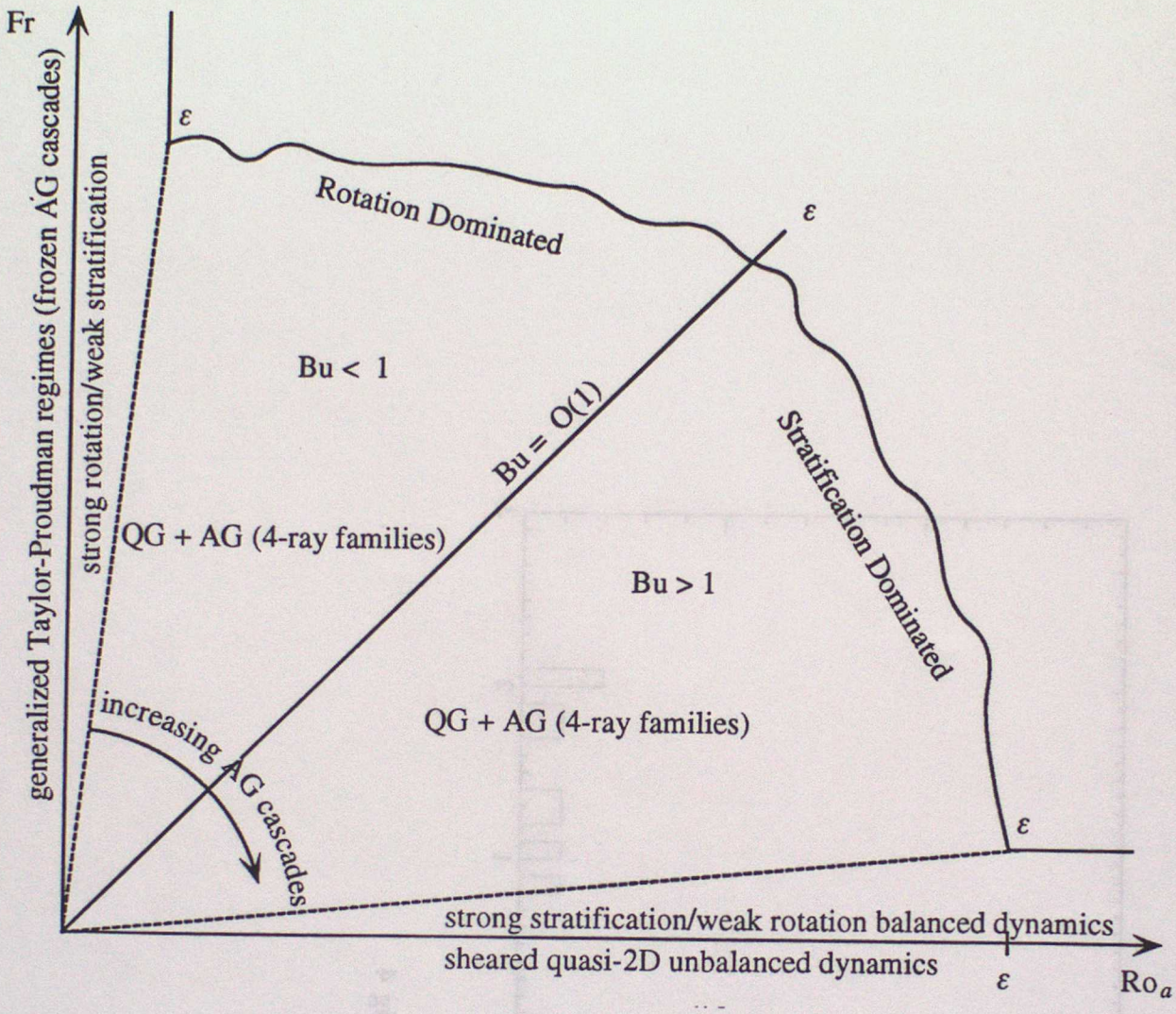

Fr
generalized Taylor-Proudman regimes (frozen AG cascades)
strong rotation/weak stratification
ε
Rotation Dominated
ε
Bu < 1
Bu = O(1)
Stratification Dominated
QG + AG (4-ray families)
Bu > 1
increasing AG cascades
QG + AG (4-ray families)
ε
strong stratification/weak rotation balanced dynamics
sheared quasi-2D unbalanced dynamics
ε
Ro$_a$

Fig 5

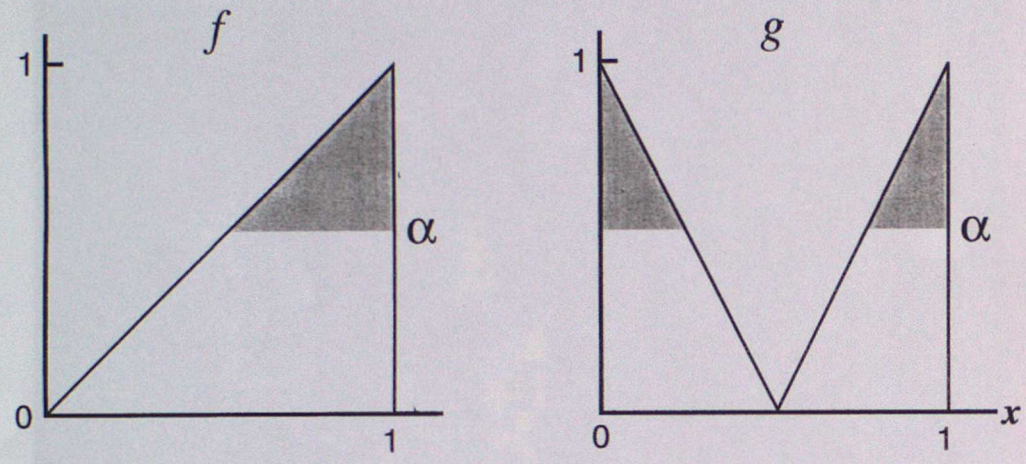


Fig 6

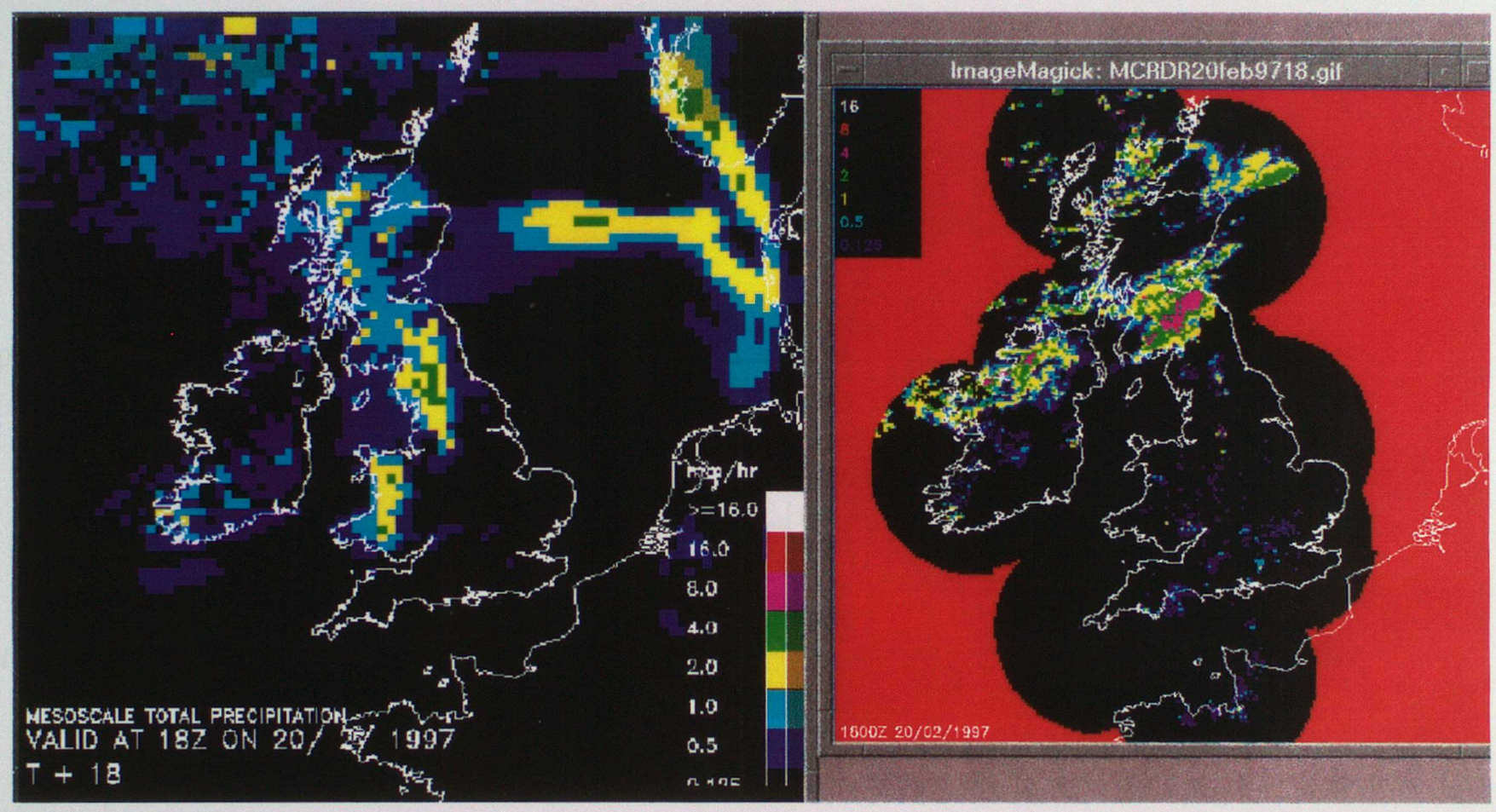

MESOSCALE TOTAL PRECIPITATION
VALID AT 18Z ON 20/ 2/ 1997
T + 18
mm/hr
>=16.0
16.0
8.0
4.0
2.0
1.0
0.5
ImageMagick: MCRDR20feb9718.gif
16
8
4
2
1
0.5
1600Z 20/02/1997

Fig 7

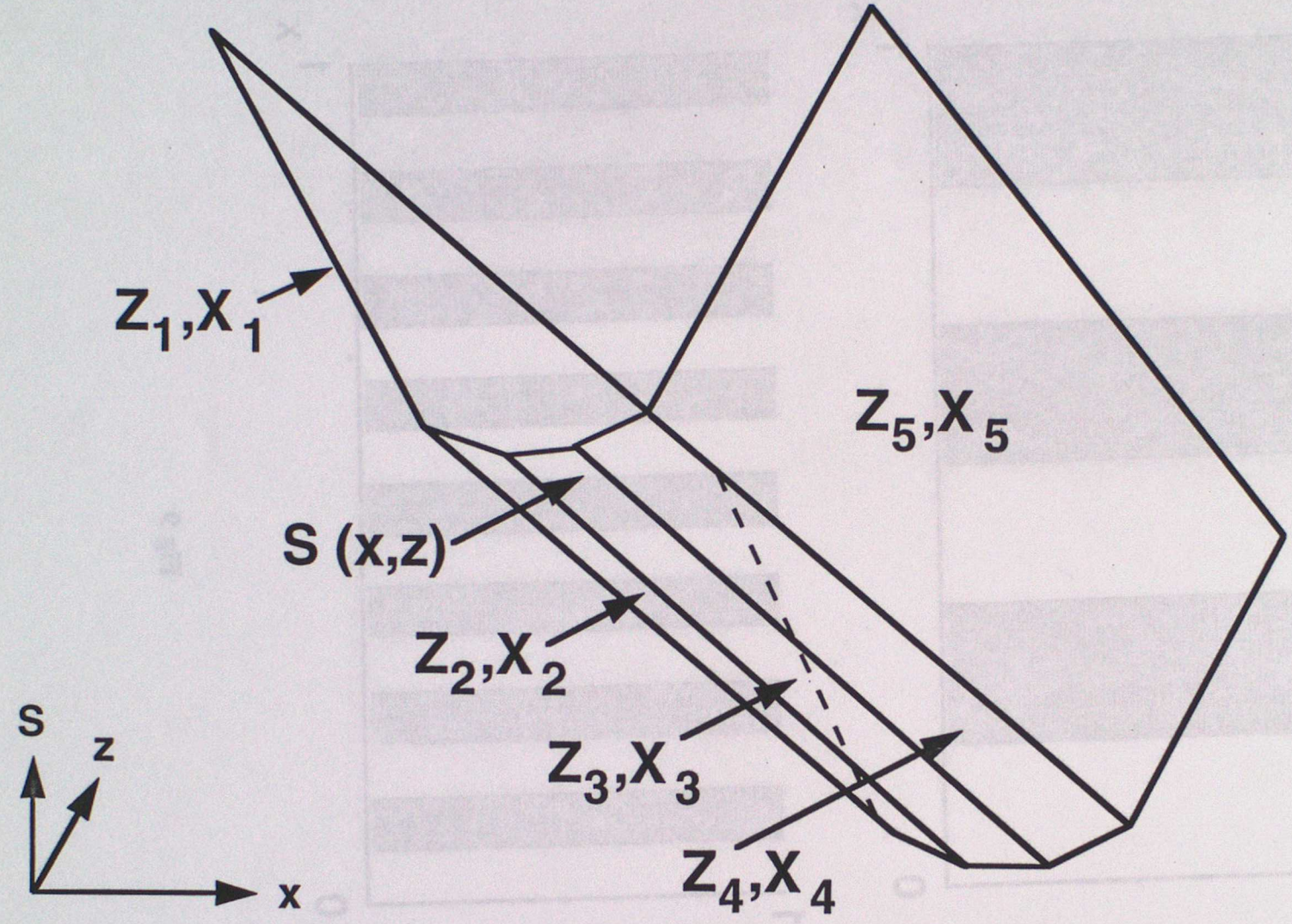


Fig 8

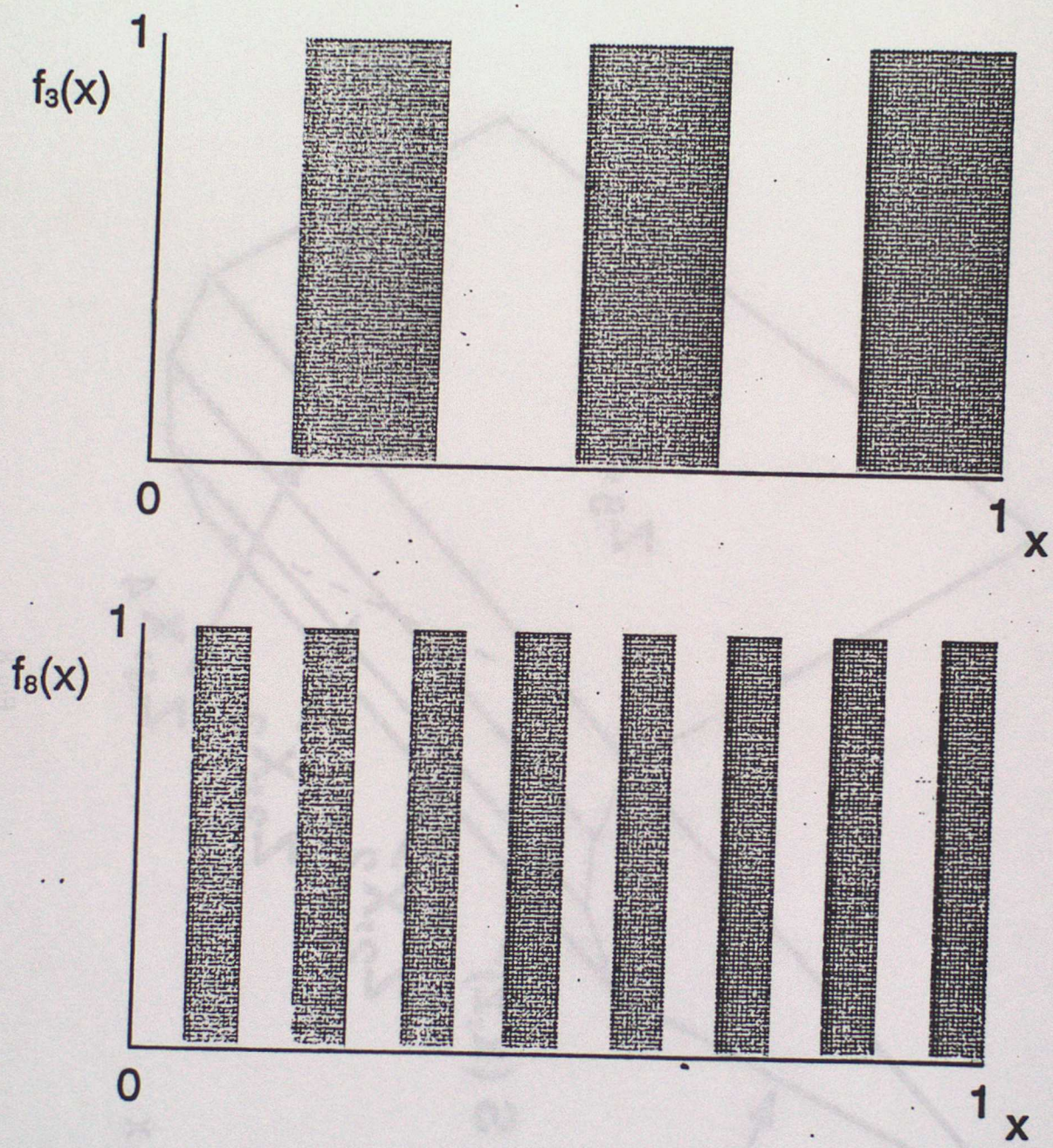
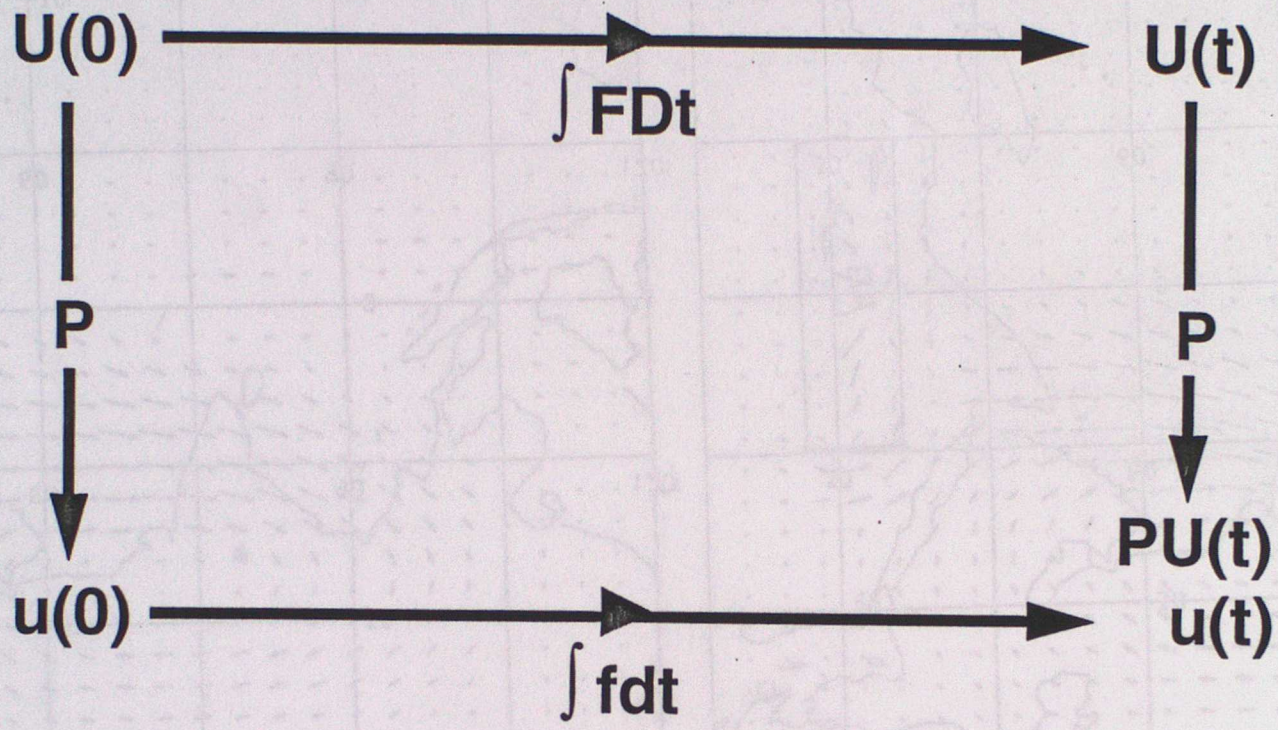


Fig 9

Let P project U onto u.
Then write forecast procedure as:



Forecast error is $||PU(t) - u(t)||$

Also interested in information loss: $||U(t) - PU(t)||$

Fig 10

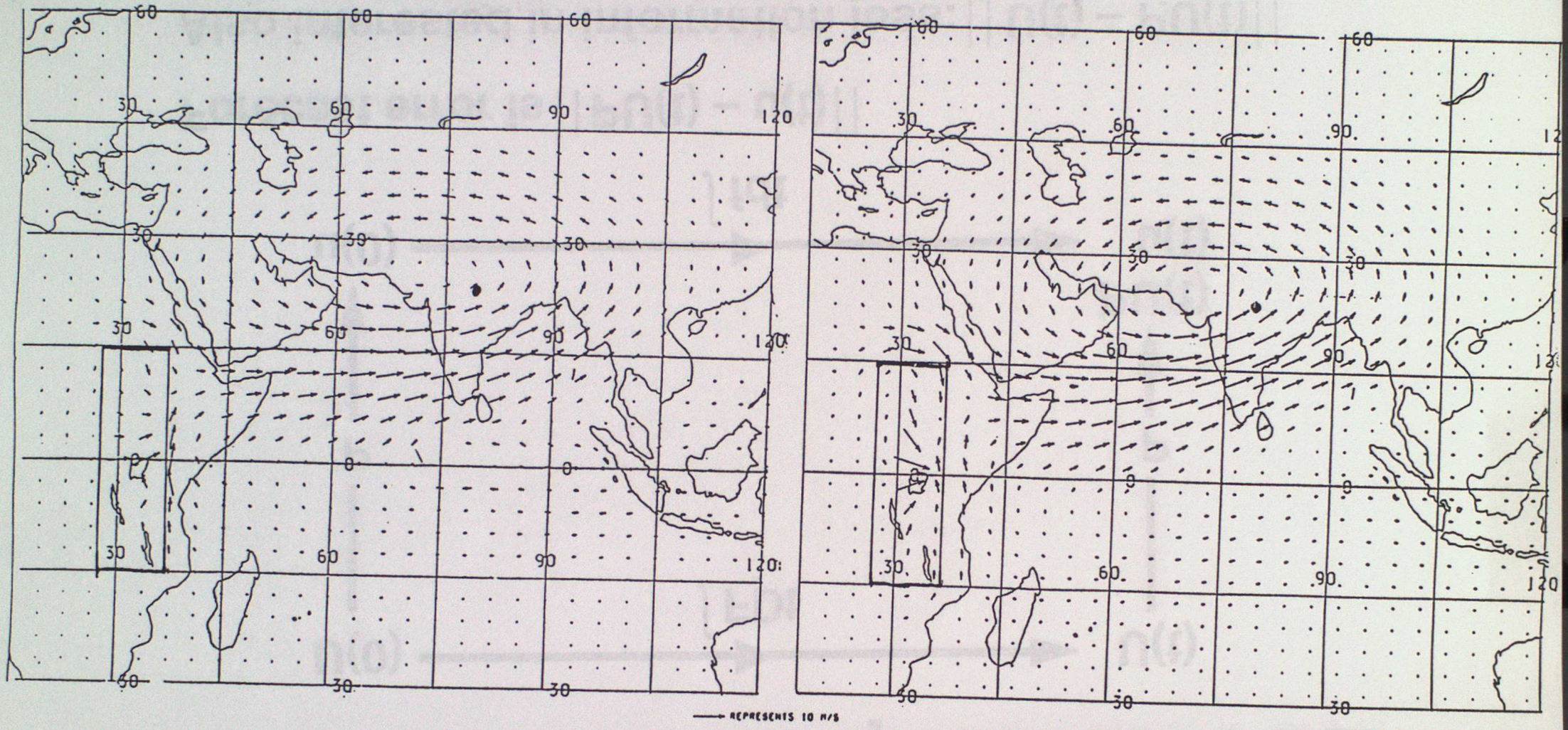


Fig 11

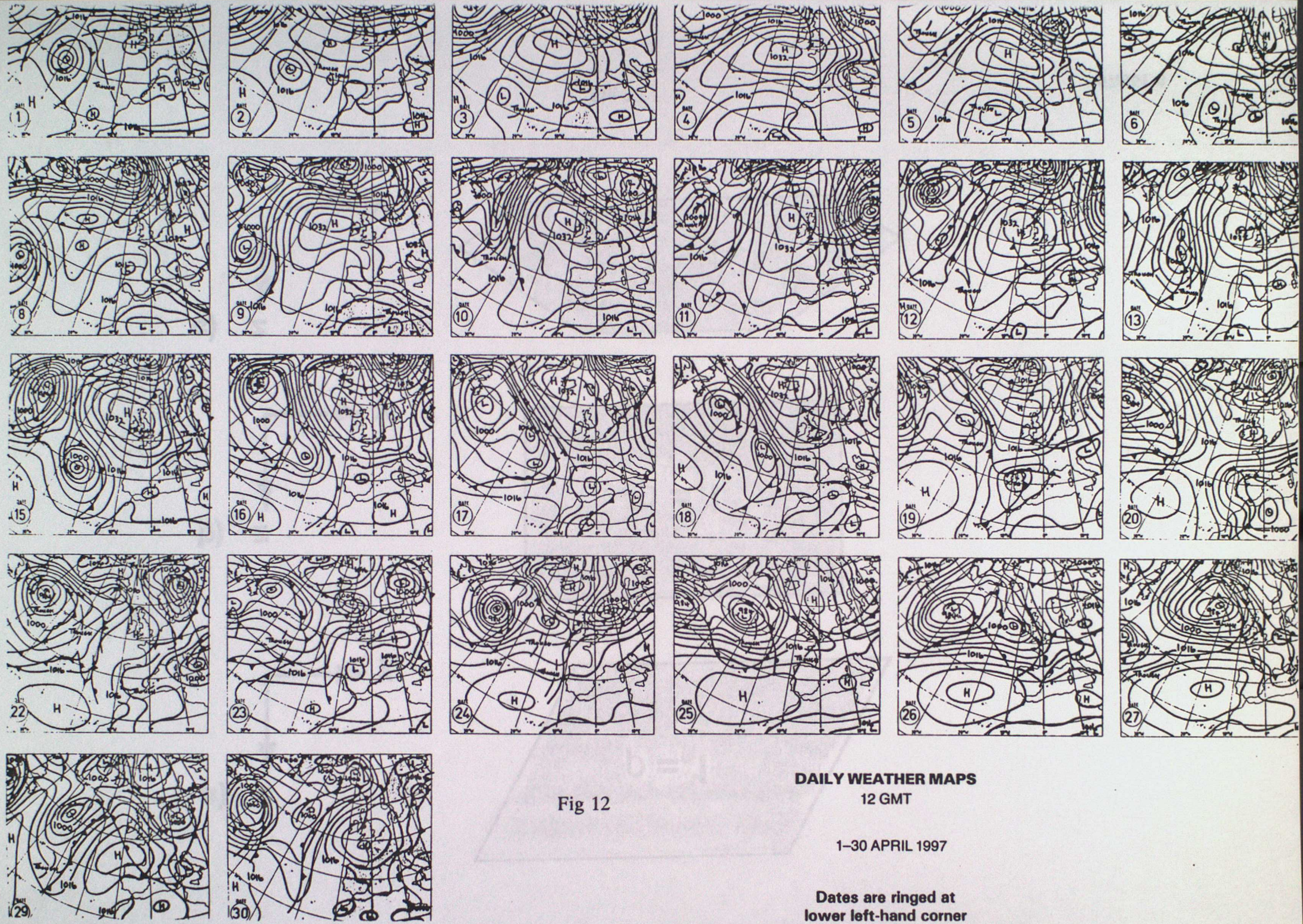


Fig 12

DAILY WEATHER MAPS
12 GMT

1–30 APRIL 1997

Dates are ringed at lower left-hand corner

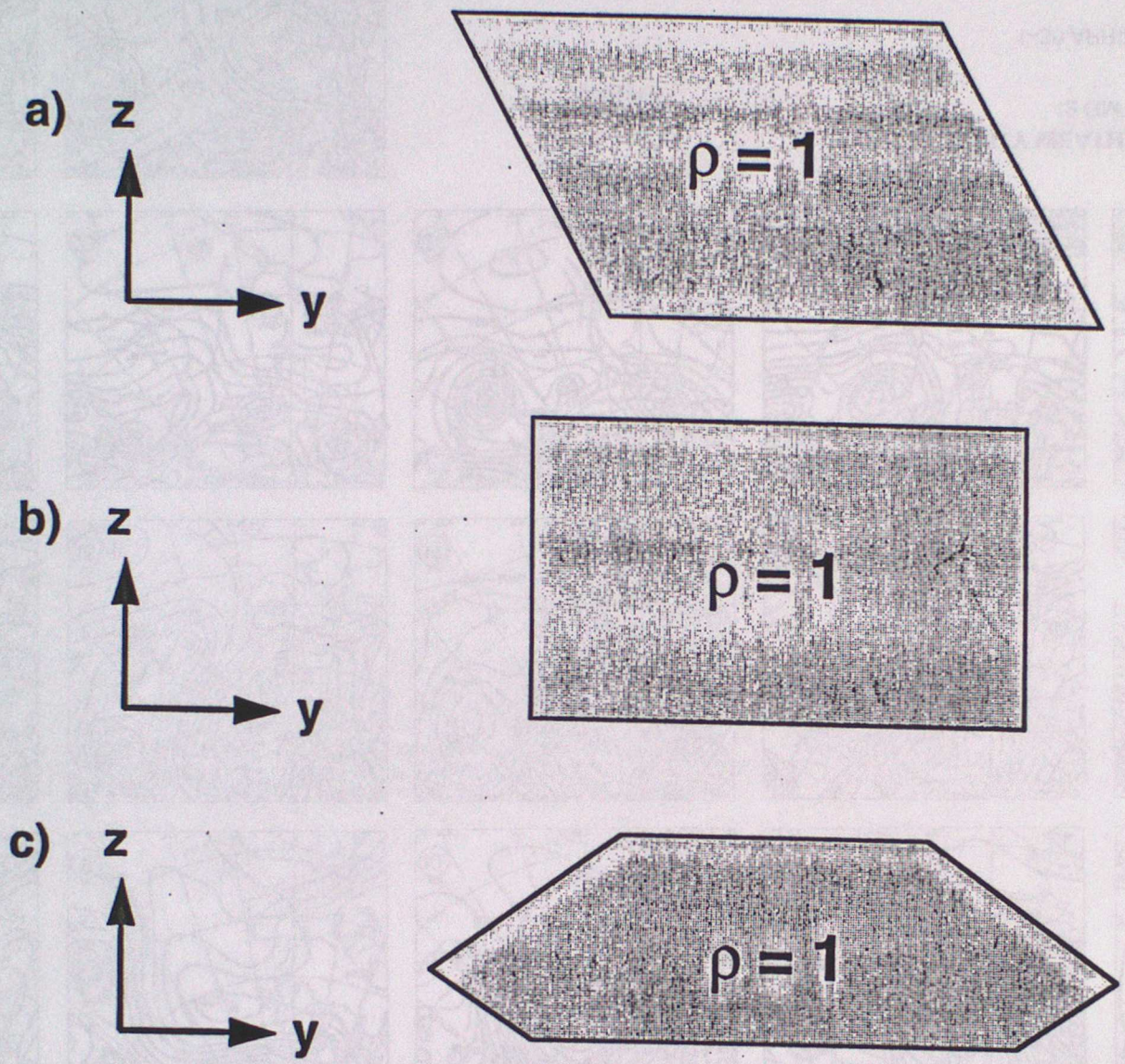


Fig 13

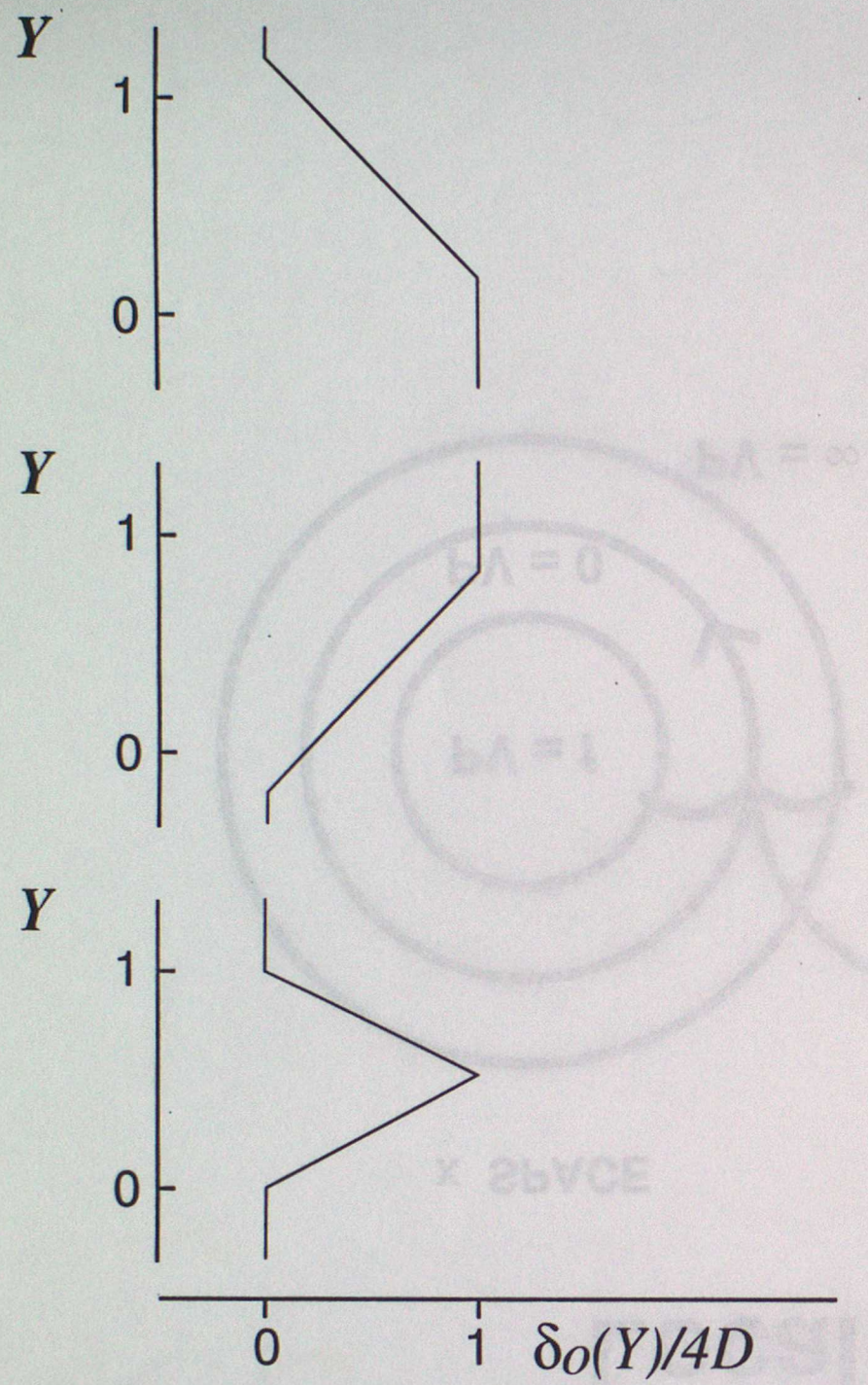


Fig 14

Local steady states

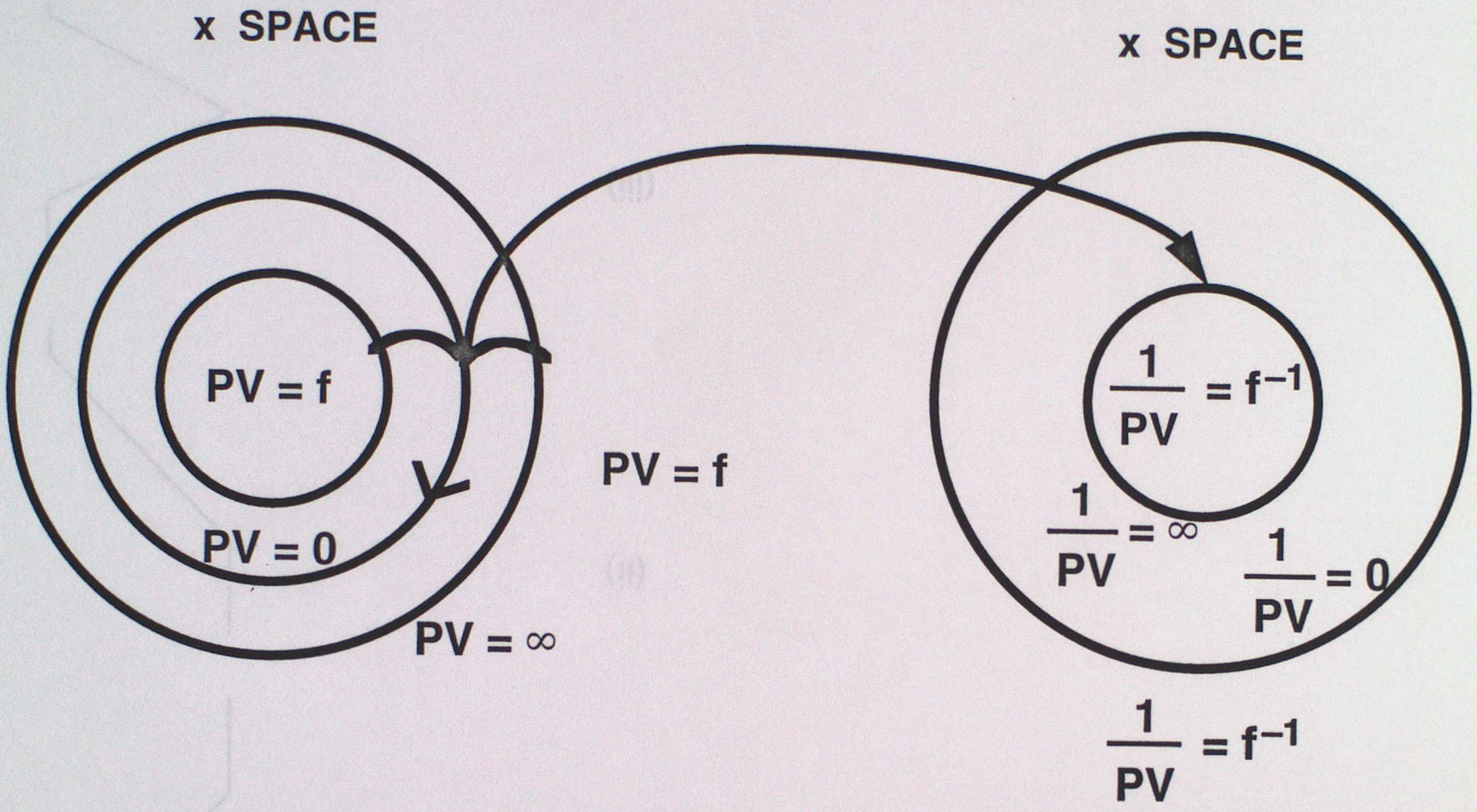


Fig 15

97/851 2

Modified semi-Lagrangian method

Preserve the volume in the departure point calculation

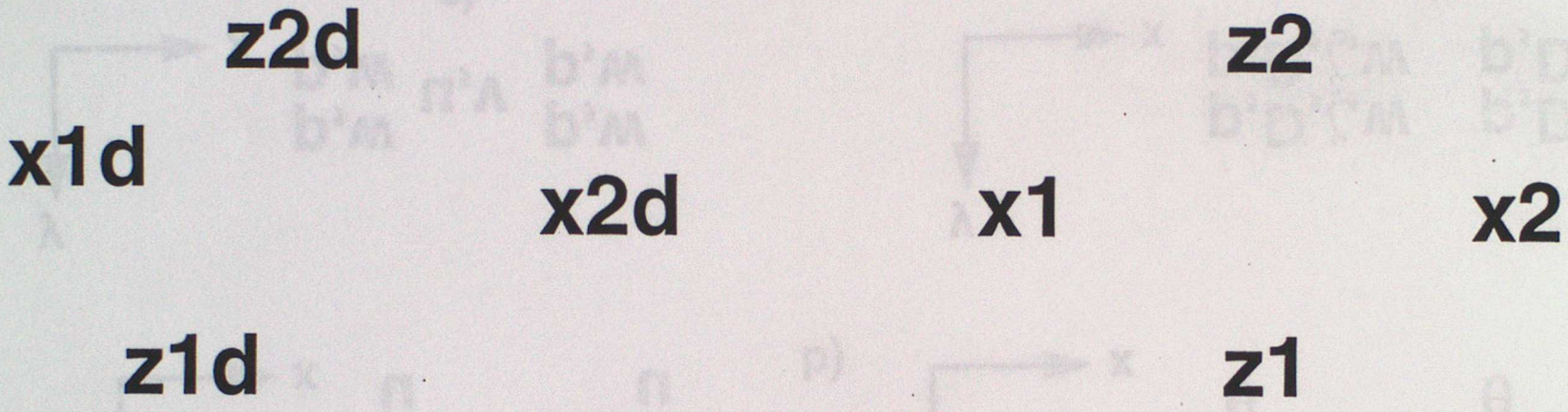


Fig 16

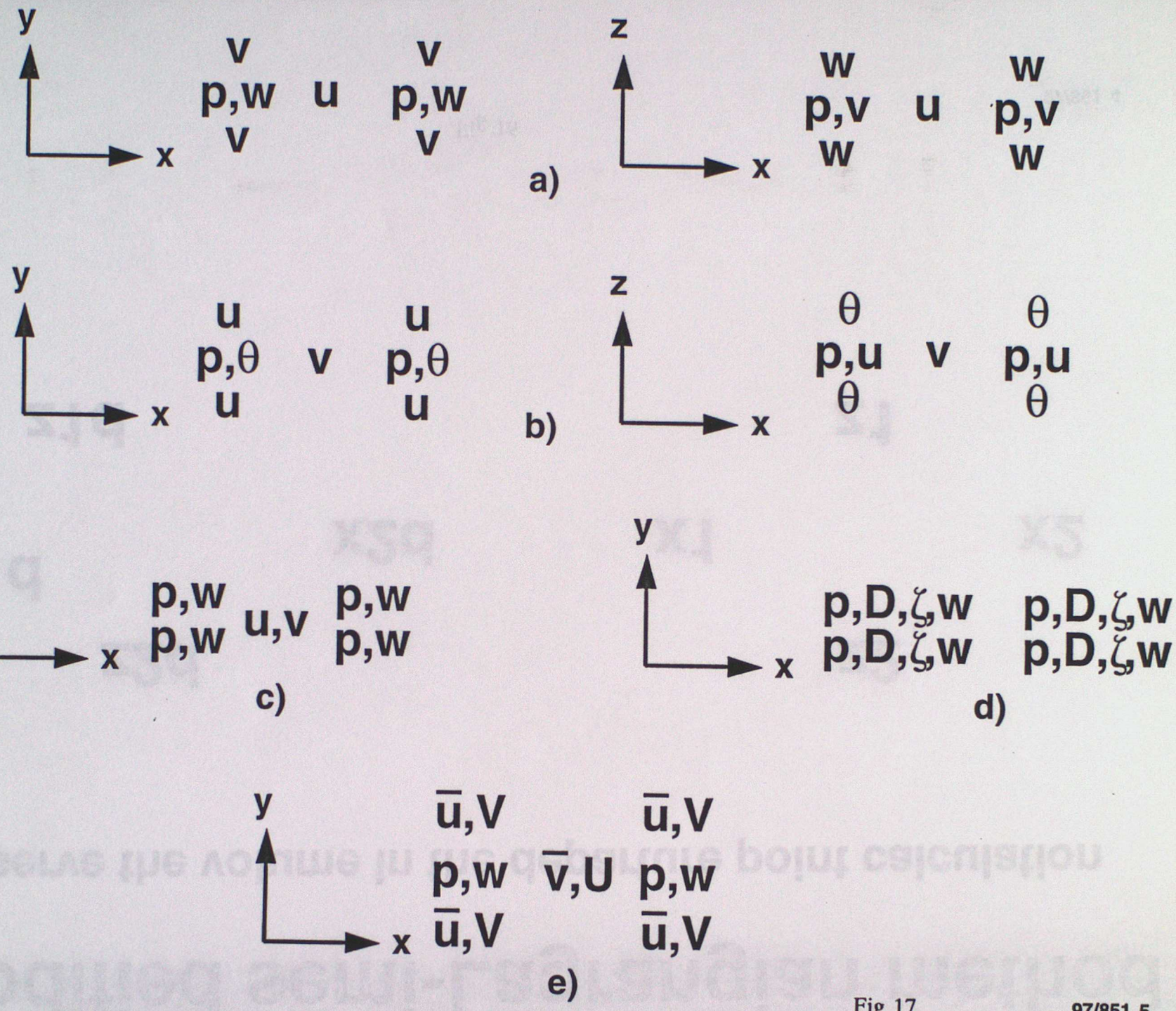

y
x
v
p,w u p,w
v
v
v
a)
z
x
w
p,v u p,v
w
w
w
y
x
u
p,θ v p,θ
u
u
u
b)
z
x
θ
p,u v p,u
θ
θ
θ
y
x
p,w
p,w u,v p,w
p,w
c)
y
x
p,D,ζ,w p,D,ζ,w
p,D,ζ,w p,D,ζ,w
d)
y
x
ū,V ū,V
p,w v̄,U p,w
ū,V ū,V
e)

Fig 17

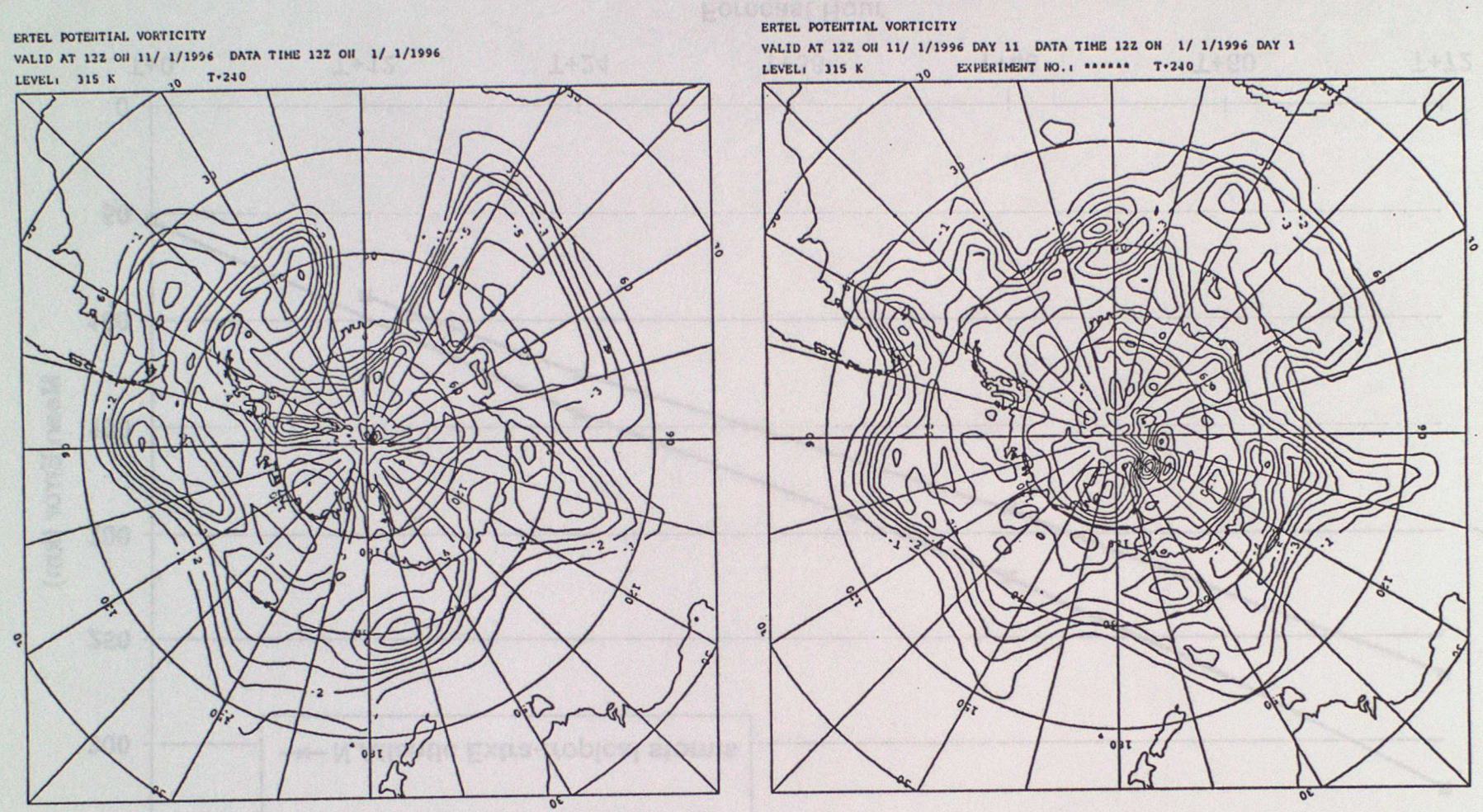

ERTEL POTENTIAL VORTICITY
VALID AT 12Z ON 11/ 1/1996 DATA TIME 12Z ON 1/ 1/1996
LEVEL: 315 K T+240
ERTEL POTENTIAL VORTICITY
VALID AT 12Z ON 11/ 1/1996 DAY 11 DATA TIME 12Z ON 1/ 1/1996 DAY 1
LEVEL: 315 K EXPERIMENT NO.: ****** T+240

Fig 18

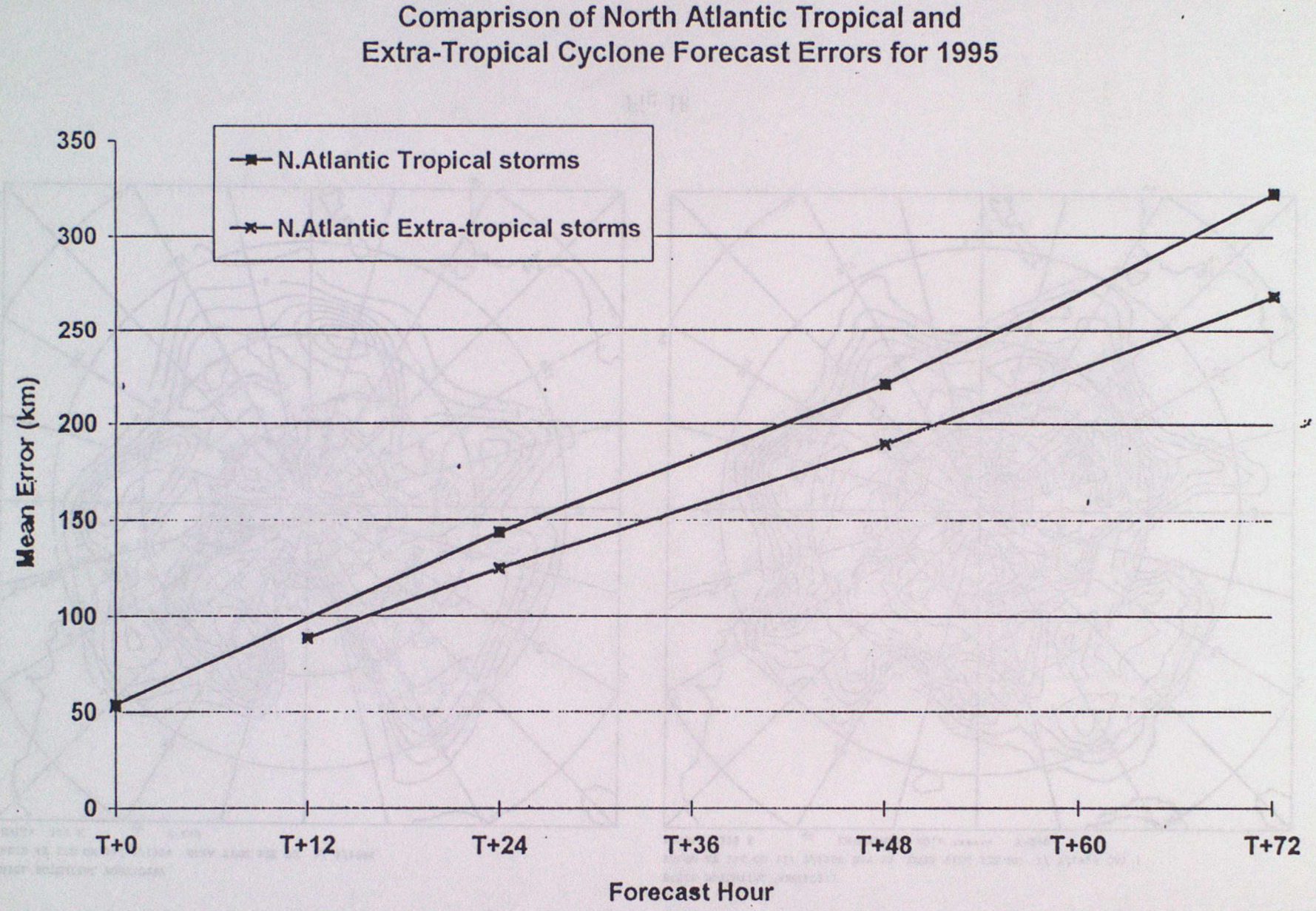

Comaprison of North Atlantic Tropical and
Extra-Tropical Cyclone Forecast Errors for 1995
N.Atlantic Tropical storms
N.Atlantic Extra-tropical storms
Mean Error (km)
350
300
250
200
150
100
50
0
T+0
T+12
T+24
T+36
T+48
T+60
T+72
Forecast Hour

Fig 19

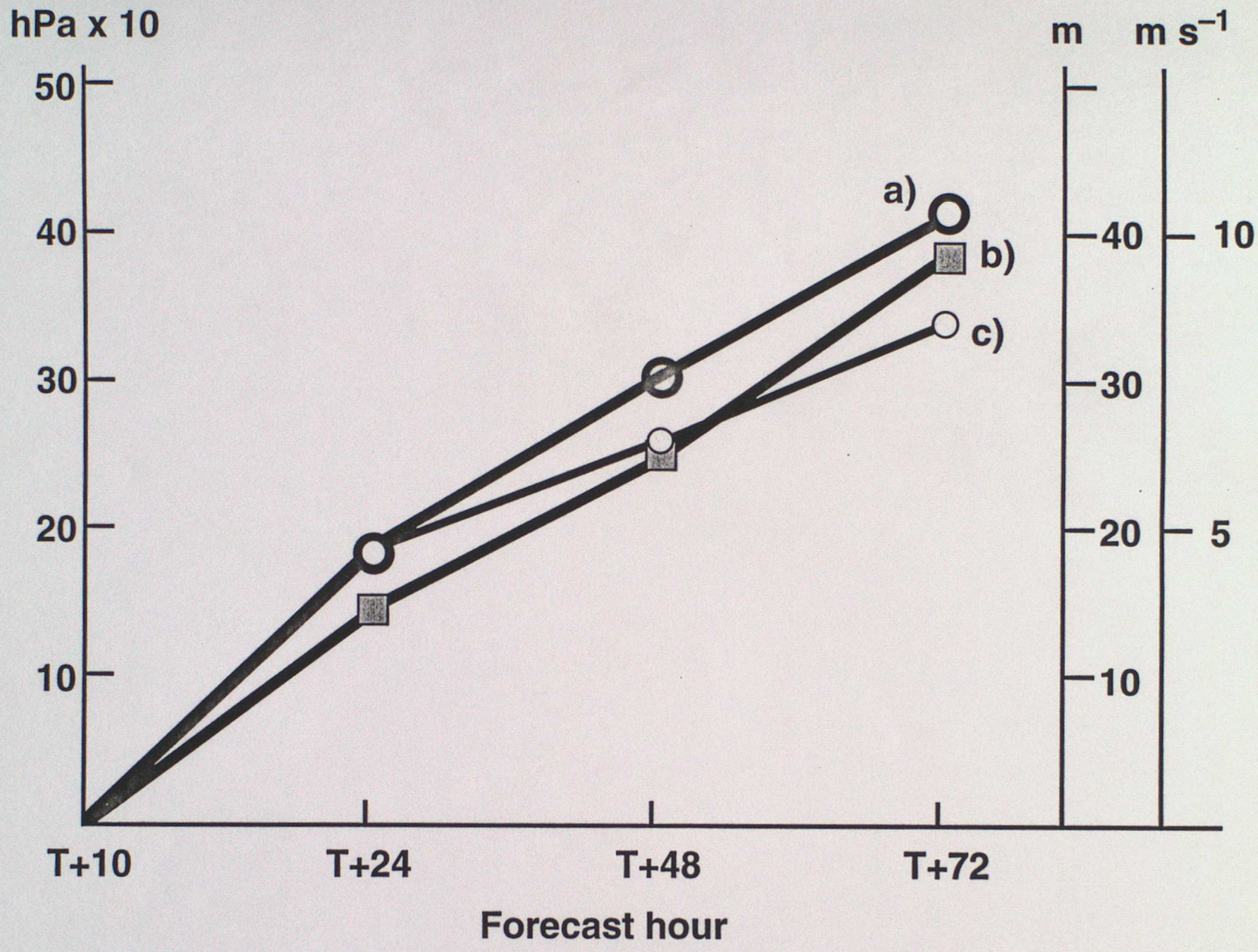


Fig 20

97/851 7